

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Search for Long-Lived Particles Decaying to Displaced Dimuons at 13 TeV and Study of Neutron-Induced Background Hits in the Muon System of the Compact Muon Solenoid

**Permalink**

<https://escholarship.org/uc/item/32s9r84f>

**Author**

Dasgupta, Abhigyan

**Publication Date**

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA  
Los Angeles

Search for Long-Lived Particles Decaying to Displaced Dimuons at 13 TeV  
and Study of Neutron-Induced Background Hits  
in the Muon System of the Compact Muon Solenoid

A dissertation submitted in partial satisfaction  
of the requirements for the degree  
Doctor of Philosophy in Physics

by

Abhigyan Dasgupta

2019

© Copyright by  
Abhigyan Dasgupta  
2019

# ABSTRACT OF THE DISSERTATION

Search for Long-Lived Particles Decaying to Displaced Dimuons at 13 TeV  
and Study of Neutron-Induced Background Hits  
in the Muon System of the Compact Muon Solenoid

by

Abhigyan Dasgupta

Doctor of Philosophy in Physics

University of California, Los Angeles, 2019

Professor Robert D Cousins, Chair

This thesis presents a search for new long-lived particles decaying to two muons in the CMS detector with pp collision data taken at  $\sqrt{s} = 13$  TeV corresponding to  $36.3 \text{ fb}^{-1}$  of integrated luminosity during Run 2 of the LHC. Such decays would appear as dimuon vertices displaced with respect to the pp interaction point. The search presented in this thesis uses only the muon system in order to probe the longest lifetimes to which the LHC experiments are sensitive. The results are interpreted in terms of a benchmark model consisting of exotic Higgs bosons decaying to long-lived scalar bosons, but are presented in an approximately model-independent way suitable for reinterpretation. No excess is observed. The appendix describes a study of neutron-induced background in CMS endcap muon chambers.

The dissertation of Abhigyan Dasgupta is approved.

David Saltzberg

Jay Hauser

Zvi Bern

Robert D Cousins, Committee Chair

University of California, Los Angeles

2019

*To my parents, who lit the way*

## TABLE OF CONTENTS

Abstract . . . . .	ii
Table of Contents . . . . .	v
List of Figures . . . . .	ix
List of Tables . . . . .	xvi
Acknowledgments . . . . .	xviii
<b>1 Introduction . . . . .</b>	<b>1</b>
<b>2 Theoretical Motivations for Long-Lived Particles . . . . .</b>	<b>3</b>
2.1 Overview of the Standard Model . . . . .	3
2.1.1 Particles, Interactions, and the Brout-Englert-Higgs Mechanism . . .	3
2.1.2 Fermi's Golden Rule for Particle Lifetimes . . . . .	6
2.2 Beyond the Standard Model . . . . .	7
2.2.1 Exotic Long-Lived Particles . . . . .	7
2.2.2 PYTHIA Configuration . . . . .	9
<b>3 The CMS Detector at the CERN LHC . . . . .</b>	<b>10</b>
3.1 The CERN LHC . . . . .	10
3.1.1 Accelerator Complex, Proton Injection Chain, and Bunch Structure .	10
3.1.2 RF Cavities and Steering Magnets . . . . .	11
3.1.3 Instantaneous and Integrated Luminosity . . . . .	14
3.2 The CMS Detector . . . . .	16
3.2.1 Introduction . . . . .	16
3.2.2 Coordinate System . . . . .	19
3.2.3 Silicon Tracker . . . . .	21

3.2.4	Electromagnetic Calorimeter . . . . .	22
3.2.5	Hadronic Calorimeter . . . . .	22
3.2.6	Solenoid Magnet . . . . .	23
3.2.7	Muon System . . . . .	24
3.2.8	Trigger System . . . . .	27
<b>4</b>	<b>Search for Long-Lived Particles Decaying to Displaced Dimuons in the CMS Muon System at <math>\sqrt{s} = 13</math> TeV . . . . .</b>	<b>31</b>
4.1	Trigger, Data, and Simulation . . . . .	31
4.1.1	Trigger . . . . .	31
4.1.2	Data Samples . . . . .	33
4.1.3	Monte Carlo Simulation Samples of Signal and Background . . . . .	33
4.1.4	Equivalent Luminosity and Event Weighting in MC Simulation . . . . .	37
4.2	Transverse Decay Length and Transverse Collinearity Angle . . . . .	38
4.3	Muon Reconstruction . . . . .	40
4.4	Event and Object Selection . . . . .	44
4.4.1	Overview . . . . .	44
4.4.2	DSA Muon Preselection . . . . .	46
4.4.3	HLT-RECO Matching . . . . .	48
4.4.4	DSA to PAT Muon Association . . . . .	48
4.4.5	DSA Muon Object Selection . . . . .	56
4.4.6	Dimuon Formation from Common Vertex Fit . . . . .	58
4.4.7	Pairing Criteria . . . . .	61
4.4.8	Dimuon Object Selection . . . . .	66
4.4.9	Dimuon Signal Selection . . . . .	66



4.4.10	Cosmic Muon Suppression . . . . .	66
4.4.11	Summary of Event and Object Selection . . . . .	69
4.5	Background Estimation . . . . .	71
4.5.1	Definition of Control Regions . . . . .	72
4.5.2	Estimation of Drell-Yan Background . . . . .	74
4.5.3	Estimation of QCD Background . . . . .	79
4.6	Cut Optimization . . . . .	83
4.6.1	Optimizing with $Z_{B_i}$ as a Figure of Merit . . . . .	84
4.6.2	Dimuon Mass Window Selection . . . . .	87
4.6.3	N-1 Plots . . . . .	89
4.7	Systematic Uncertainties . . . . .	89
4.7.1	Luminosity Uncertainty . . . . .	94
4.7.2	Background Uncertainties . . . . .	94
4.7.3	Signal Uncertainties . . . . .	97
4.8	Results . . . . .	102
4.8.1	Computing Upper Limits with the $CL_s$ Criterion . . . . .	102
4.8.2	Technical Details of Using the Higgs Combine Tool . . . . .	104
4.8.3	Events Passing the Full Selection Criteria . . . . .	106
4.8.4	Upper Limits on the Cross Section Times Branching Ratio . . . . .	106
4.8.5	Robustness of Upper Limits to Systematic Uncertainties . . . . .	107
<b>5</b>	<b>Summary and Concluding Remarks . . . . .</b>	<b>121</b>

<b>A Neutron-Induced Background Hits in the CMS Endcap Muon System and Implications for the HL-LHC . . . . .</b>	<b>123</b>
Foreword . . . . .	123
Abstract of the Appendix . . . . .	123
A.1 Introduction . . . . .	124
A.2 Cathode Strip Chambers and Trigger Electronics in CMS . . . . .	126
A.3 GEANT4 MC Simulation for Neutron Studies . . . . .	130
GEANT4 MC Simulation Setup . . . . .	130
Results from MC Simulation . . . . .	132
A.4 Selection and Isolation of Neutron-Induced Hits in CMS Data . . . . .	136
A.5 Results and Comparison of CMS Data with MC Simulation . . . . .	142
Computation of Neutron-Induced Hit Rates . . . . .	142
Results . . . . .	155
Neutron Hit Patterns . . . . .	158
A.6 Muon Test Beam Studies at GIF++ . . . . .	158
The GIF++ Experimental Setup . . . . .	158
LCT Efficiency and RecHit Displacement at GIF++ . . . . .	160
A.7 Charge per Hit Studies at GIF++ and in CMS Data and Simulation . . . . .	166
Computation of charge/hit from CMS data . . . . .	166
Computation of charge/hit from GIF++ data . . . . .	180
Computation of charge/hit from CMS simulation . . . . .	180
Comparison of charge/hit . . . . .	180
A.8 Acknowledgments of the Appendix . . . . .	183
<b>References . . . . .</b>	<b>184</b>

## LIST OF FIGURES

3.1	Overview of the CERN accelerator complex. . . . .	12
3.2	Summary of the injection chain of the LHC. . . . .	12
3.3	Diagram of the bunch places of an LHC proton beam showing filled and empty places. . . . .	13
3.4	Phase space diagram of an RF bucket and the elliptical boundary of a proton bunch within the bucket. . . . .	15
3.5	Diagram of cross section of an LHC steering dipole magnet and depiction of the magnetic field lines in the magnets. . . . .	15
3.6	Total pp integrated luminosity recorded by CMS during Run 1 and Run 2 by year and month. . . . .	17
3.7	Transverse slice of the CMS detector illustrating the basic structure of CMS and the shapes of the tracks and energy deposits formed by the five general categories of particles directly detected by CMS. . . . .	19
3.8	Schematic diagram of one quadrant of CMS in $r$ - $z$ . . . . .	20
3.9	Radius of curvature from sagitta. . . . .	24
3.10	Principle of operation of a drift tube cell. . . . .	26
3.11	Muon $p_T$ resolution as a function of muon $p_T$ for the barrel and endcap using the tracking system only, muon system only, and combining both. . . . .	28
3.12	Architecture of the L1 trigger system, depicting the local, regional, and global components of the muon and calorimeter triggers. . . . .	30
4.1	Diagram of key variables ( $L_{xy}$ and $ \Delta\Phi $ ) used in the analysis. . . . .	39
4.2	DSA and RSA reconstruction efficiency as a function of generated $L_{xy}$ . . . . .	43
4.3	DSA and RSA $p_T$ resolution for generated events within acceptance . . . . .	43
4.4	DSA muon $p_T$ resolution with and without selections on $N_{\text{hits}}$ and $\sigma_{p_T}/p_T$ . . . .	47

4.5	DSA muon charge difference with and without selections on $N_{\text{hits}}$ and $\sigma_{p_T}/p_T$ . . .	47
4.6	Flowchart depicting the technical details of the DSA to PAT association procedure	52
4.7	Histogram of dimuon $L_{xy}/\sigma_{L_{xy}}$ in simulated background samples before and after the PAT association procedure. . . . .	54
4.8	Histogram of dimuon $L_{xy}$ in simulated background samples before and after the PAT association procedure. . . . .	54
4.9	Graph of the number of dimuons after the PAT association divided by the number of dimuons before the PAT association, as a function of generated $L_{xy}$ , in all $2\mu$ signal samples combined. . . . .	55
4.10	Efficiency for the common vertex fit to converge as a function of generated $L_{xy}$ for generated events within acceptance and histograms of $\sigma_{L_{xy}}$ normalized to unit area for $L_{xy} < 320$ cm and $L_{xy} > 320$ cm. . . . .	60
4.11	Diagram illustrating the application of pairing criteria to dimuons in the case of four selected muons. . . . .	63
4.12	Flowchart depicting the technical details of the pairing criteria procedure. . . .	64
4.13	Efficiency for the pairing criteria to correctly choose reconstructed dimuons as a function of generated $L_{xy}$ for $2\mu$ and $4\mu$ signal samples. . . . .	65
4.14	Efficiency for the pairing criteria to correctly choose reconstructed dimuons. . .	65
4.15	Display of an event in data consistent with multiple cosmic muons from a single atmospheric shower in a pp collision event. . . . .	68
4.16	Histogram of $ \Delta\Phi $ for all $2\mu$ simulated signal samples combined, showing a strong peak at zero. . . . .	72
4.17	Histograms of $ \Delta\Phi $ in data and in MC simulation for the DY control regions with DSA to PAT association reversed, for all values of $ \Delta\Phi $ , for all values of $L_{xy}/\sigma_{L_{xy}}$ and for $L_{xy}/\sigma_{L_{xy}} > 6$ . . . . .	75
4.18	Diagrams depicting the effect of dimuon orientation on possible values of $L_{xy}$ and $\sigma_{L_{xy}}$ . . . . .	77

4.19	Histograms of $ \Delta\Phi $ for sequentially increasing cut values of $L_{xy}/\sigma_{L_{xy}}$ in data and Drell-Yan simulation. . . . .	78
4.20	Histogram of $ \Delta\Phi $ for sequentially increasing cut values of $L_{xy}/\sigma_{L_{xy}}$ for QCD-like events in data. . . . .	80
4.21	Histograms of $m_{\mu\mu}$ and $\Delta R(\mu\mu)$ for DY-like events in data for sequentially increasing cut values of $L_{xy}/\sigma_{L_{xy}}$ . . . . .	82
4.22	Histograms of $m_{\mu\mu}$ and $\Delta R(\mu\mu)$ for QCD-like events in data for sequentially increasing cut values of $L_{xy}/\sigma_{L_{xy}}$ . . . . .	82
4.23	Example of $Z_{Bi}$ -based cut optimization of $L_{xy}/\sigma_{L_{xy}}$ . . . . .	86
4.24	Reconstructed invariant dimuon mass distributions for $2\mu$ signal samples, long with fitted Gaussian curves for each distribution. . . . .	88
4.25	Histograms of events passing the full selection except for the muon $p_T$ cut in $2\mu$ signal and data. . . . .	90
4.26	Histograms of events passing the full selection except for the distance of closest approach cut in $2\mu$ signal and data. . . . .	90
4.27	Histograms of events passing the full selection except for the barrel $N(\text{DT hits})$ cut in $2\mu$ signal and data. . . . .	91
4.28	Histograms of events passing the full selection except for the track $\chi^2/\text{dof}$ cut in $2\mu$ signal and data. . . . .	91
4.29	Histograms of events passing the full selection except for the $\cos\alpha$ and $N(\text{parallel pairs})$ cuts in $2\mu$ signal and data. . . . .	92
4.30	Histograms of events passing the full selection except for the $\cos\alpha$ and $N(\text{parallel pairs})$ cuts in $2\mu$ signal and data. . . . .	92
4.31	Histograms of events passing the full selection except for the vertex $\chi^2$ cut in $2\mu$ signal and data. . . . .	93
4.32	Histograms of events passing the full selection except for the $L_{xy}/\sigma_{L_{xy}}$ cut in $2\mu$ signal and data. . . . .	93

4.33	Histograms of ${}^0_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$ and ${}^{\pi}_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$ in data, with the latter binwise scaled by Drell-Yan and QCD transfer factors, demonstrating agreement and validating the method of obtaining background estimates in the signal region with transfer factors from control regions. . . . .	96
4.34	Histograms of the number of true primary vertices in 2016 data and in $2\mu$ signal, and graph of their ratios. . . . .	100
4.35	Display of $\rho$ - $\phi$ and $\rho$ - $z$ views of event 1675672448 in 2016 data, in run 274969, lumi section 944. . . . .	109
4.36	Display of $\rho$ - $\phi$ and $\rho$ - $z$ views of event 801995838 in 2016 data, in run 276935, lumi section 522. . . . .	110
4.37	Display of $\rho$ - $\phi$ and $\rho$ - $z$ views of event 59821031 in 2016 data, in run 276940, lumi section 97. . . . .	111
4.38	Display of $\rho$ - $\phi$ and $\rho$ - $z$ views of event 6168006 in 2016 data, in run 276947, lumi section 4. . . . .	112
4.39	Display of $\rho$ - $\phi$ and $\rho$ - $z$ views of event 203474554 in 2016 data, in run 277168, lumi section 114. . . . .	113
4.40	Display of $\rho$ - $\phi$ and $\rho$ - $z$ views of event 188965749 in 2016 data, in run 278017, lumi section 169. . . . .	114
4.41	Display of $\rho$ - $\phi$ and $\rho$ - $z$ views of event 635154343 in 2016 data, in run 278308, lumi section 390. . . . .	115
4.42	Display of $\rho$ - $\phi$ and $\rho$ - $z$ views of event 676135274 in 2016 data, in run 281797, lumi section 501. . . . .	116
4.43	95% CL upper limits on $\sigma(\text{H} \rightarrow \text{XX})B(\text{X} \rightarrow \mu\mu)$ for $m_{\text{H}} = 1000$ GeV for various values of $m_{\text{X}}$ . . . . .	117
4.44	95% CL upper limits on $\sigma(\text{H} \rightarrow \text{XX})B(\text{X} \rightarrow \mu\mu)$ for $m_{\text{H}} = 400$ GeV for various values of $m_{\text{X}}$ . . . . .	118

4.45	95% CL upper limits on $\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)$ for $m_H = 200$ GeV for various values of $m_X$ . . . . .	119
4.46	95% CL upper limits on $\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)$ for $m_H = 125$ GeV for various values of $m_X$ . . . . .	119
A.1	Diagram and principle of operation of a CSC endcap muon chamber in CMS. . .	127
A.2	Diagram of the CSC electronics system in CMS. . . . .	129
A.3	Histogram of neutron kinetic energy just before nuclear capture, before and after enabling thermal treatment of detector nuclei. . . . .	131
A.4	Final energy of simulated neutron vs. the time since pp collision giving rise to simulated detector hit, for hits in CSCs. . . . .	134
A.5	$r$ - $z$ view of CMS cavern showing the locations of the specific neutron captures that led to SimHits in the CSCs. . . . .	135
A.6	Stacked histogram of the final energy of simulated photons at the time they produce simulated electrons that lead to simulated detector hits. . . . .	135
A.7	Opposite half selection for neutron-induced hits in CSCs. . . . .	137
A.8	Histogram of bunch crossing number in a particular CMS data taking period during LHC proton-proton collisions in 2016, showing the gaps of various sizes in the LHC bunch structure. . . . .	138
A.9	2D histogram of number of anode wire hits from CSCs in the inner ring of the first station (ME1/1) of the CMS endcap muon system. . . . .	140
A.10	Plot of $N_{pp}^{CMS}/t$ vs. instantaneous luminosity. . . . .	148
A.11	Plot of instantaneous luminosity vs. mean pileup. . . . .	149
A.12	Plot of hits per time per area as a function of luminosity for ME1/1 chambers. .	153
A.13	Plot of hits per time per area as a function of luminosity for ME2/1 chambers. .	154

A.14 Histogram of neutron capture induced anode wire hits per time per area for CMS data and for MC simulation for ME1/1 chambers for a reference luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . . . . .	156
A.15 Histogram of neutron capture induced anode wire hits per time per area for CMS data and for MC simulation for ME2/1 chambers for a reference luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . . . . .	157
A.16 Distribution of candidate neutron capture induced patterns in 2016 CMS data and CMS simulation. . . . .	159
A.17 Plot of LCT efficiency divided by the efficiency at minimum HV current for ME1/1 and ME2/1 chambers at GIF++. . . . .	161
A.18 Display of an event collected during a muon test beam at the CERN Gamma Irradiation Facility (GIF++) with a CSC from ME2/1. . . . .	162
A.19 Display of the same event, showing offline reconstructed detector responses (Re-cHits). . . . .	163
A.20 Stacked plot of fraction of offline reconstructed muon segments, categorized by the number of reconstructed hits used to form them, for the ME1/1 and ME2/1 at GIF++. . . . .	165
A.21 Plot of anode HV current vs. LHC luminosity with a two-parameter linear fit for three ME1/1 channels. . . . .	167
A.22 Histogram of offset-corrected reference HV currents from ME1/1 HV channels. .	169
A.23 Plot of ME1/1 wire group $N_{\text{hits}}^{\text{CMS}}/t$ vs. LHC luminosity with a one-parameter linear fit for hits that occur during LHC bunch gaps and during pp collisions. .	170
A.24 Plot of current vs. luminosity for ME+2/1/04 layer 5 for three HV Segments with a two-parameter linear fit. . . . .	172
A.25 Plot of ME2/1 wire group $N_{\text{hits}}^{\text{CMS}}/t$ vs. LHC luminosity for hits that occur during LHC bunch gaps and during pp collisions. . . . .	173
A.26 Plot of anode wire group $N_{\text{hits}}^{\text{CMS}}(d)/t$ that occur during pp collisions in ME2/1. .	174



A.27 Plot of HV current vs. hits/s for an ME2/1 layer with smooth transitions in current and a jump in HV current. . . . .	177
A.28 Histogram of values of charge/hit in fC in ME2/1 for LHC Fill 5394. . . . .	178
A.29 Plot of anode HV current vs. hits/s at GIF++ for ME1/1 and ME2/1 HV segments.	179
A.30 Histogram of the avalanche charge per simulated wire hit for for ME1/1 and non-ME1/1 chambers in CMS simulation. . . . .	181

## LIST OF TABLES

2.1	Fermion (spin-1/2) content of the Standard Model. . . . .	5
2.2	Boson (spin-1 and spin-0) content of the Standard Model. . . . .	5
3.1	LHC nominal design beam parameters. . . . .	16
4.1	CMS datasets used by the analysis, including dataset names, the “Muon Physics” run ranges as certified by CMS for physics analysis, and integrated luminosities. . . . .	33
4.2	Simulated $H \rightarrow XX$ signal samples used by the analysis, for several combinations of BSM Higgs mass ( $m_H$ ), long-lived particle mass ( $m_X$ ), and $c\tau$ . . . . .	35
4.3	CMS simulation samples used by the analysis for each background process, along with the total production cross section and the number of generated events. . . . .	36
4.4	Summary of full selection, organized into event, muon, and dimuon requirements. . . . .	70
4.5	Event counts for control regions for estimating Drell-Yan background. . . . .	79
4.6	Event counts for control regions for estimating QCD background. . . . .	83
4.7	Dimuon invariant mass window selections for each value of generated long-lived particle mass $m_X$ . . . . .	87
4.8	Transfer factors $\tau_{\text{QCD}}$ for estimating QCD background, along with their system- atic uncertainties, given in dimuon mass windows. . . . .	95
4.9	Summary of systematic uncertainties in this analysis for signal and background estimates. . . . .	101
4.10	List of events passing the full selection in 2016 data, with their run, lumi section, and event numbers, the values of $m_{\mu\mu}$ , $ \Delta\Phi $ , and $L_{xy}/\sigma_{L_{xy}}$ , and the associated event display. . . . .	108
4.11	Number of opposite-sign and same-sign events, expected background events, and observed events. . . . .	108

4.12	95% CL upper limits on $\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)$ for the 33 sets of signal parameters $(m_H, m_X, c\tau)$ in $2\mu$ simulated signal samples. . . . .	120
A.1	Enumeration of time windows used for each region listed by digi readout time bins and bunch places after the gap. . . . .	141
A.2	Values of wire group hit fraction and hit rate in each ME2/1 HV segment. . . .	175
A.3	Values of charge/hit in ME1/1 and each HV segment of ME2/1 in units of fC/hit.	182

## ACKNOWLEDGMENTS

First, I thank my CMS collaborators and colleagues, without whom this work would not have been possible.

The work on displaced dimuons was performed in collaboration with UCLA senior researcher Slava Valuev and UCLA postdoctoral scholar Alice Florent under the supervision of my advisor, Professor Robert Cousins, along with UCLA colleagues Jay Hauser and William Nash, and Institute of High Energy Physics Vienna colleagues including Wolfgang Adam, Alberto Escalante, Suchita Kulkarni, Ivan Mikulec, Sebastian Templ, and Claudia Wulz. We also thank our colleagues and conveners in CMS who contributed valuable knowledge and discussions, including Juliette Alimena, Paul Lujan, Ivan Mikulec, Niki Saoulidou, and Adish Vartak, and other UCLA professors Michalis Bachtis, Zvi Bern, and David Saltzberg.

The work on neutron background studies described in the appendix was performed in full collaboration with fellow UCLA graduate student Christian Schnaible under the supervision of our common advisor, Professor Robert Cousins. Together, Christian Schnaible and I wrote a detailed draft of internal CMS Detector Note DN-2017/019, reproduced below. Professor Cousins also provided editing assistance, particularly in Section A.5. I presented this work in a poster session at the 2017 European Physical Society Conference on High-Energy Physics (EPS-HEP), and wrote a conference note for the Proceedings. Christian Schnaible presented this work in a talk at the the 2017 American Physical Society Meeting of the Division of Particles & Fields (APS-DPF).

We also thank everyone on the CSC GIF++ team, the CMS UCLA group, and everyone whose work we have built on or partially reproduced and who have contributed valuable knowledge and advice, including Cameron Bravo, Tim Cox, Alice Florent, Jay Hauser, Misha Ignatenko, Vladimir Ivantchenko, Evaldas Juska, Alexey Kamenev, Andrey Korytov, Katerina Kuznetsova, Armando Lanaro, Paul Lujan, Alex Madorsky, Nick McColl, Hualin Mei, Yuriy Pakhotin, Vladimir Palchik, Victor Perelygin, Jian Wang, Wells Wulsin, and Piet Verwilligen.

This thesis includes material that is based upon work supported by the U.S. Department of Energy under Award Number DE-SC0009937.

I also thank my friends who have personally supported me over the years. You have listened to me, provided me with understanding and caring emotional support that has kept me healthy, sane, and happy, and been the keystones of many experiences that have filled my life with vibrancy—weekends at Shawn’s, Christmas parties, Friendsgivings, trivia at Mr. Pickwick’s Pub, movie nights, S’Beer nights, Ferney markets, D&D campaigns, French lunches, coffee and chess afternoons, and countless lunches, dinners, fireworks, BBQs, lake-side swims, wine tastings, hikes, and trips across sixteen European countries. I cannot thank you enough for my life of the past four years; the least I can do is attempt to mention some of your names here.

My high school friends including Adam Alonzo, Emily McGreevy Alonzo, Danielle Caponpon, Andy Chen, Jeff Cruz, Derek Dosland, Madeline Foss, Brandie Hanson, Krystal Lin, Arohi Sharma, and Dustin Tygret. My Berkeley friends including Spandan Bandyopadhyay, Memo Carpenter, Alexey Drobizhev, Nitin Egbert, Harshil Goel, Alex Latshaw, Kelsey Oliver-Mallory, Michela Paganini, and Ana Vasudevan. My internet friends including Katherine Rybski and Mahathi Chavali.

My UCLA friends including Pauline Arriaga, Cameron Bravo, Gary Chen, Krishna Choudhury, Eric Cotner, Eve Emmanouilidou, Rhyann Ghosh, MaryAnn Kontonikolas, Alex Latshaw, Nathan Majernik, Ana Martynova, David Perkins, Seth Pree, Jeff Robertson, Chris Schnaible, Grace Stees, Joe Swearngin, Shawn Tang, and Sam Witte.

And my CERN friends including Juliette Alimena, Reyer Band, Chris Brainerd, Cameron Bravo, Alex Brown, Kira Burt, Sarah Charley, Grace Cummings, Andrés Delannoy, Pieter Everaerts, Juliana Froggatt, Amin Ghiasi, Usama Hussain, Rachel Hyneman, Jan Kiesler, Tony Lefeld, Bing Liu, Kenneth Long, Paul Lujan, Chris Martin, Christine McLean, Vivian Nguyen, Michela Paganini, Ashley Parker, Michael Planer, Brendan Regnery, Elodie Resseguie, Francesca Ricci-Tam, Titas Roy, Lesley Smith, Nick Smith, Eric Takasugi, Dylan Teague, Santona Tuli, Ana Vasudevan, Matthias Weber, Nate Woods, and Brent Yates.

I also thank my family, without whom I would not be here. My parents, Sanjukta Dasgupta and Alope Dasgupta, have encouraged me, supported me, instilled me with values and culture, and provided me with everything to succeed. No amount of thanks can accurately capture their impact on my life. My cousins Soham Dasgupta, Twinkle Ghosh, and Ritwika Kanungoe, who have kept me connected with a distant dream. My adopted family Sonia Das and Mihir Das, whose powerful support have formed a backbone of my life. And everyone else in my family who have showered me with love.

Last, but not least, I thank my advisor, Professor Robert Cousins. I have learned a great deal from following Bob's guidance and listening to his advice, and his values and knowledge, often delivered as concise aphorisms, have found a permanent place in my memory. His high standards have pushed me to learn and accomplish more than I could have expected. And he has been generous and organized, ensuring minimal worries about logistical issues and allowing me to focus on performing the research contained within this thesis. For all this and more I am grateful, and I consider myself lucky to have been his student.

## VITA

2009–2013 Undergraduate Research Assistant, CONCEPT Laboratory, UC Berkeley.

2013 B.A. (Physics) and B.A. (Mathematics), UC Berkeley.

2013–2015 Teaching Assistant, Physics Department, UCLA.

2014 M.S. (Physics), UCLA.

2015–present Graduate Student Researcher, Physics Department, UCLA.

# CHAPTER 1

## Introduction

*Quantum field theory arose out of our need to describe the ephemeral nature of life.*

—A. Zee, *Quantum Field Theory in a Nutshell*

Elementary particle physics is a field of scientific inquiry that attempts to answer one of humanity’s deepest and most fundamental questions: what are we made of? The unifying power of its underlying theory has been nothing short of spectacular, describing three of the four fundamental forces of nature and all known fundamental particles in a single breath of elegant mathematics. This theory is known as the Standard Model, one of the crown jewels of twentieth century particle physics, and so successful has it been in fact that to date, nearly all measurements performed by four decades of ever increasingly sophisticated experiments have been harmonious with its predictions.

Nevertheless, the Standard Model is believed to be incomplete for a number of reasons, ranging from unfulfilled aesthetic guiding principles to unexplained experimentally observed phenomena such as gravity, dark matter, and neutrino mass. A renowned achievement of the field is the discovery of the Higgs boson in 2012, and its subsequent confirmation (as predicted by the Standard Model) as the particle formed from interactions giving all known particles their mass. Yet since then, no new particles, heralds of an exciting landscape of unexplored phenomena, have been confirmed. The quest continues.

High-energy proton-proton (pp) collisions surrounded by a network of particle detectors provide a lens with which to study the Standard Model and to search for hints of what lies beyond. This thesis presents a search for new particles with long lifetimes, in a parameter space not yet fully explored by the large, general-purpose experiments based at the Large



Hadron Collider (LHC) at CERN. The analysis is performed using Compact Muon Solenoid (CMS) data taken during Run 2 of the LHC, and is heir to two CMS analyses [1, 2] performed with data taken during Run 1 of the LHC. A similar analysis was performed by the ATLAS collaboration using data taken during Run 2 of the LHC [3].

No analysis searching for a faint hint of new physics over a sea of background effects could be performed without a keen understanding of the behavior of the experimental apparatus, and so this thesis also presents a contribution to that understanding: a measurement of background hits induced by neutrons in the cathode strip chamber muon detectors found in the endcaps of CMS, with data taken both by the CMS experiment and at the CERN Gamma Irradiation Facility (GIF++) located near the CERN Super Proton Synchrotron. Potential implications of this background for the performance of the High-Luminosity LHC (HL-LHC) are also discussed.

The thesis is organized as follows. Chapter 2 presents theoretical motivations for the existence of these particles, as well as arguing for the discovery sensitivity of these experiments. Chapter 3 is an overview of the CMS experimental apparatus within the accelerator complex of the LHC. Chapter 4 describes the search for new long-lived particles decaying to two muons, using CMS data taken in 2016 at a center-of-mass energy of 13 TeV corresponding to  $36.3 \text{ fb}^{-1}$  of integrated luminosity. This analysis searches for vertices characteristically displaced from the proton-proton collision point, signaling the potential decay of a long-lived particle. Chapter 5 summarizes the analysis and proposes extensions to it for further study. Appendix A presents the study of neutron-induced background hits in the endcap muon system of CMS, and studies performed with muon test beam data taken by similar muon chambers at GIF++.

## CHAPTER 2

### Theoretical Motivations for Long-Lived Particles

*What is the pattern, or the meaning, or the why? It does not do harm to the mystery to know a little about it. For far more marvelous is the truth. . .*

—R. Feynman, *The Feynman Lectures on Physics, Volume I*

This chapter begins with a brief overview of the Standard Model (SM), a highly successful theory of the electromagnetic, weak, and strong fundamental forces describing the kinematics of and interactions between all known elementary particles. Despite its success, the SM is believed to be incomplete, both because there exist experimentally observed phenomena not explained by the SM, and because there are a few theoretical loose ends and unfulfilled guiding aesthetic principles. This chapter then continues with a brief motivation for the existence of as-of-yet unobserved long-lived particles, illustrating models in which they can arise and showing the existence of experimental sensitivity for detecting such new particles. In this chapter and throughout the rest of this thesis, units are used in which  $\hbar = c = 1$ , except when using the quantity  $c\tau$ , i.e.  $c$  times the mean proper lifetime of a particle.

## 2.1 Overview of the Standard Model

### 2.1.1 Particles, Interactions, and the Brout-Englert-Higgs Mechanism

The SM is (mathematically) formalized as a gauge quantum field theory described by a Lagrangian density composed of quantum fields [4, 5, 6, 7, 8]. The defining characteristic of this Lagrangian is its invariance under local transformations of the fields under the gauge group

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \quad (2.1)$$

By virtue of Noether’s theorem, continuous symmetries of the Lagrangian correspond to conserved currents; here,  $\mathrm{SU}(3)_C$  corresponds to conservation of color charge;  $\mathrm{SU}(2)_L$  corresponds to weak isospin; and  $\mathrm{U}(1)_Y$  corresponds to weak hypercharge. Tables 2.1–2.2 enumerate the fundamental fermions and bosons of the SM, along with their quantum numbers and gauge symmetries.

Terms of the Lagrangian generally fall into one of three categories: kinetic terms involving derivatives, mass terms that are second-order in the fields, and interaction terms that are third-order and higher in the fields describing interactions between fields. The kinetic and mass terms together are called the free Lagrangian, since they describe a free field not interacting with anything else.

The SM Lagrangian begins with free terms for the fermions described by Dirac Lagrangians and free terms for the bosons associated with gauge groups described by Proca Lagrangians, all initially without any mass terms. Interaction terms between fermions and bosons are generated from the free SM Lagrangian by requiring that it be invariant under local gauge transformations, a requirement achieved by replacing all derivatives in the fermion kinetic terms with an appropriate “covariant derivative” that varies with choice of gauge. The terms of the covariant derivatives are such that the experimentally observed interactions between fermions and bosons and their strengths are reproduced and incorporated into the theory; for this reason the fundamental bosons are also referred to as gauge bosons.

All SM fermions and the W and Z gauge bosons are observed to have a mass, but the previous treatment assumes all particles to be massless. Adding appropriate mass terms would result in the Lagrangian losing its local gauge invariance. Therefore, in the SM, particles acquire mass terms by interacting with a scalar  $\mathrm{SU}(2)_L$  doublet  $\phi$  [8]. Below an energy threshold (the electroweak symmetry breaking scale), its potential,

$$V = -\mu^2\phi^2 + \lambda\phi^4 \quad (2.2)$$

Generation			$C$	$T$	$T_3$	$Y$	$Q$
	I	II	III				
Quarks	$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	r, g, b	1/2	+1/2 -1/2	+1/3 -1/3
	$u_R$	$c_R$	$t_R$	r, g, b	0	0	+4/3 -1/3
	$d_R$	$s_R$	$b_R$				
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	0	1/2	+1/2 -1/2	0 -1
	$e_R$	$\mu_R$	$\tau_R$	0	0	0	-2 -1

Table 2.1: Fermion (spin-1/2) content of the Standard Model. Fermions are presented in their Cabbibo-Kobayashi-Maskawa-rotated flavor eigenstates, linear combinations of their mass eigenstates. Fermions are also presented decomposed into their chiral components; subscripts  $L$  and  $R$  refer to the handedness of the chiral component. Quantum numbers given are color charge  $C$ , weak isospin  $T$ , the third component of weak isospin  $T_3$ , the weak hypercharge  $Y = 2(Q - T_3)$ , and the electric charge  $Q$  [5, 7].

Field	Gauge Group	Number	Name	Symbol	Gauge Group	Number
$G$	$SU(3)_C$	8	Gluon	$g$	$SU(3)_C$	8
$W$	$SU(2)_L$	3	W	$W^\pm$		2
			Z	$Z$		1
$B$	$U(1)_Y$	1	Photon	$\gamma$	$U(1)_{\text{em}}$	1
$\phi$		4	Higgs	$h$		1

Table 2.2: Boson (spin-1 and spin-0) content of the Standard Model (left) before and (right) after electroweak symmetry breaking, along with their number and their associated gauge group. Interactions with the field  $\phi$  break the  $SU(3)_C \times SU(2)_L \times U(1)_Y$  gauge symmetry into  $SU(3)_C \times U(1)_{\text{em}}$ , with three of the four bosons generated by  $SU(2)_L \times U(1)_Y$  acquiring a mass by having absorbed three of the four Goldstone bosons associated with  $\phi$  [4, 5, 8].

exhibits a nonzero vacuum expectation value, resulting in the continuous symmetry associated with gauge transformations of  $\phi$  to be spontaneously broken. The electroweak symmetry breaking scale sets the scale of the masses of the W and Z bosons.

Spontaneous electroweak symmetry breaking results in a reshuffling of fundamental bosons; the mechanism by which it is accomplished is known as the Brout-Englert-Higgs (BEH) mechanism. After symmetry breaking and the BEH mechanism, the massless gauge bosons have mixed to form the experimentally observed mass eigenstates (the massive W and Z bosons and the massless photon), the massless fermions have acquired masses, and a new scalar particle, the spin-0 Higgs boson, remains.

### 2.1.2 Fermi's Golden Rule for Particle Lifetimes

A Lagrangian in a quantum field theory prescribes a set of Feynman rules for performing perturbative calculations of scattering and decay amplitudes and hence of the experimentally measurable physical quantities of cross sections and decay rates. A fundamental result of quantum field theory, then, is the relativistic version of Fermi's Golden Rule for particle decays [4, 9]. The partial decay rate  $d\Gamma$  of a particle of mass  $M$  into some number of particles is

$$d\Gamma = \frac{(2\pi)^4}{2M} |\mathcal{M}|^2 d\Phi \quad (2.3)$$

which is a product of two factors that are functions of the outgoing momenta: the matrix element (or amplitude) for the process  $\mathcal{M}$ , which describes the dynamics of the interactions, is computed from the appropriate Feynman calculus, and thus depends on gauge couplings; and the phase space factor (or density of final states)  $d\Phi$ , which describes the kinematics of the interactions, subject to momentum and energy conservation, and thus depends on the masses.

For a two-body decay, the integral over the phase space is kinematically determined, independent of the functional form of the amplitude.

$$\Gamma = \frac{|\mathbf{p}|}{16\pi M^2} |\mathcal{M}|^2 \quad (2.4)$$

where  $|\mathbf{p}|$  is the outgoing three-momentum of one of the decay products; if they have the same mass  $m$ ,

$$|\mathbf{p}| = \frac{1}{2M} \sqrt{M^4 - 4M^2 m^2} \quad (2.5)$$

Then the lifetime of the particle is the reciprocal of the decay rate:

$$\tau = \frac{1}{\Gamma} \quad (2.6)$$

## 2.2 Beyond the Standard Model

### 2.2.1 Exotic Long-Lived Particles

Although the SM has been highly successful in providing accurate, precise, and experimentally verified predictions of measurements of all known phenomena related to the strong and electroweak forces, it nonetheless is believed to be incomplete. One set of reasons for this belief is unexplained experimentally observed phenomena: renormalizable and locally gauge invariant quantum field theories incorporating gravity (via reconciliation with general relativity), neutrino oscillations (via neutrino masses), and dark matter (via new, massive, weakly interacting particles) all require extensions to the SM. As none of the many such extensions to the SM (detailed discussions of which are beyond the scope of this thesis) have as of yet found experimental support, searches for new particles beyond the SM (BSM) continue.

Comprehensive searches at the Large Hadron Collider (LHC) considering a wide variety of experimentally observable final states have yielded no observations of new particles with short lifetimes consistent with decays to electrons, muons, or jets of hadrons appearing to originate from the collision point of two proton beams. However, new particles with longer lifetimes may still exist. Such particles are not predicted by the SM, but from Equation 2.4, they may find a theoretical grounding: with appropriate choices of values for the particle masses and their gauge couplings, the phase space factors or the amplitudes may be small. This would result in low decay rate, translating into a long particle lifetime. Such exotic long-lived particles could decay to SM particles, such as muons; this would allow their production

cross sections to be probed at the LHC via a characteristic signature of a dimuon vertex formed at macroscopic distances from the proton-proton interaction point.

A possible production mechanism for these particles could be decays of the SM Higgs boson or exotic Higgs-like bosons [10]. Consider a (second) Higgs doublet  $H$  and a scalar boson  $X$ , with the potential

$$V = -\mu^2 H^2 + \lambda H^4 + M^2 X^2 + \kappa X^4 + \zeta X^2 H^2 + aX + bX^3 + cXH^2 \quad (2.7)$$

It is assumed that after electroweak symmetry breaking,  $m_H > 2m_X$ , so that the decay  $H \rightarrow XX$  can occur. For small values of  $a$ ,  $b$ , and  $c$ ,  $X$  and  $H$  may mix slightly, so that the mass eigenstate is  $X + \epsilon H$  with  $\epsilon$  small. Then the  $X$  may decay via any of the  $H$  decays, with  $\Gamma_X = \epsilon^2 \Gamma_H$ ; the  $X$  branching fractions are those of the  $H$ , but its lifetime may be quite long.

This model is merely an example of a possible extension to the SM with kinematics that are possible to study at the LHC. Other models – such as a hidden abelian Higgs model (HAHM) with dark photons decaying to muons [11] – yield similar kinematics. The intent is simply to illustrate the existence of self-consistent, flexible models with free parameters that extend the SM and motivate searches for long-lived particles.

The results of the experimental search for displaced dimuon vertices presented in Chapter 4 are interpreted in terms of a benchmark model approximating this BSM Higgs model (described in Section 2.2.2) and is used for comparison with previous results. Results are given as 95% confidence level (CL) upper limits on the cross section for production of the long-lived scalar particles via BSM Higgs decays ( $\sigma(H \rightarrow XX)$ ) times the branching fraction of the long-lived scalar particles decaying to two muons ( $B(X \rightarrow \mu\mu)$ ), for a variety of values of the BSM Higgs mass  $m_H$  and the long-lived particle mass  $m_X$ , as a function of  $c$  times the long-lived particle mean proper lifetime,  $c\tau$ . If  $H$  is the SM Higgs boson with a mass of 125 GeV, then the 95% CL upper limit on its branching fraction to BSM particles as computed by a combination of data taken at center-of-mass energy  $\sqrt{s} = 7$  and 8 TeV in 2011 and 2012 by the CMS and ATLAS collaborations is 34% [12], and the results can be used to exclude a range of long-lived particle lifetimes. For non-SM Higgs bosons, the couplings are free and unknown. Although the results are interpreted for this specific model,

parametrizing them in terms of  $m_H$ ,  $m_X$ , and  $c\tau$  presents the results in a manner that is approximately model independent, and the results can be reinterpreted to derive limits on a variety of other models.

### 2.2.2 PYTHIA Configuration

For the purposes of interpreting the results of the search presented in this thesis and to compare the results to other analyses, PYTHIA 8 [13, 14] is used to simulate high-energy collision events for a benchmark model with a BSM Higgs boson decaying to two long-lived scalar bosons, at least one of which decays to two muons. This benchmark model approximates the BSM Higgs model described in Section 2.2.1 as follows:

- Two long-lived scalar bosons  $X$  and  $X'$  of various masses  $m_X$  are defined:
  - Both particles have the same generated mass  $m_X$ , Breit-Wigner distribution width of 0.01 GeV, and nominal  $c\tau$  in millimeters.
  - Both particles have zero spin, zero electric charge, and zero color charge.
  - $X$  decays exclusively via  $X \rightarrow \mu^+ \mu^-$ .
  - $X'$  decays via  $X' \rightarrow q\bar{q}$  with 33% branching fractions to each of the three light flavors:  $u$ ,  $d$ , and  $s$ .
- BSM Higgs bosons  $H$  of various masses  $m_H$  are defined:
  - They are produced (exclusively) via gluon-gluon fusion.
  - They are produced with a Breit-Wigner distribution width of 2.7% of  $m_H$ .
  - They decay exclusively either via  $H \rightarrow XX'$  (yielding a  $2\mu$  final state) or via  $H \rightarrow XX$  (yielding a  $4\mu$  final state). This is simply a means to produce two sets of signal events, one in which both long-lived particles decay to muons and another in which just one of them decays to muons.



## CHAPTER 3

### The CMS Detector at the CERN LHC

*The expectations of life depend upon diligence; the mechanic that would perfect his work must first sharpen his tools.*

—Confucius, *Analects of Confucius*

The Compact Muon Solenoid (CMS) detector is a general-purpose detector for studying the physics of fundamental particles produced by proton-proton (pp) and heavy ion collisions at the Large Hadron Collider (LHC) at CERN. Two beams of protons circle the 27.6 km circumference of the LHC in opposite directions and collide at various locations along it; one of these locations is the CMS detector.

#### 3.1 The CERN LHC

The LHC is a two-ring circular hadron collider designed to collide protons at a center-of-mass energy of  $\sqrt{s} = 14$  TeV and at a design instantaneous luminosity of  $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  [15]. Run 2 of the LHC began in 2015 with its superconducting dipole magnets operating such that the corresponding center-of-mass energy is  $\sqrt{s} = 13$  TeV [16].

##### 3.1.1 Accelerator Complex, Proton Injection Chain, and Bunch Structure

Figure 3.1 is a diagram of the CERN accelerator complex. The LHC tunnel has eight arcs and eight straight sections. Each straight section can serve as a location for experiments, but only four are used as such. Beam crossings occur at four of these points: the locations of the four largest experiments at the LHC. The two high-luminosity, general-purpose experiments,

CMS and ATLAS, are located at Point 5 and Point 1, respectively. The two lower luminosity, more special purpose experiments, ALICE and LHCb, are located at Point 2 and Point 8, respectively. Yellow dots indicate these four experiments in Figure 3.1, each with their own large detectors in underground caverns receiving collisions from the LHC.

Figure 3.2 is a diagram summarizing the injection chain by which protons are accelerated in the LHC. Protons begin at Linac2, a linear accelerator, and are subsequently injected into the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), the Super Proton Synchrotron (SPS), and finally the LHC. Filling the LHC requires 12 cycles of the SPS and 3–4 cycles of the PS, yielding a total LHC filling time of approximately 4 minutes per beam.

Proton beams at the LHC are not continuous streams of protons, but rather organized into high-intensity bunches spaced 25 ns apart, corresponding to a crossing frequency of 40 MHz. The LHC has 3564 bunch places, of which up to 2808 are filled with protons in colliding bunches. Consecutive bunches of protons occur in trains of filled bunches separated by gaps. Proton-proton collisions occur when bunches of protons cross, called bunch crossings. Figure 3.3 is a diagram showing the 3564 bunch places, each represented by a square, and the pattern of filled bunches, represented by squares filled in blue.

### 3.1.2 RF Cavities and Steering Magnets

Protons are accelerated from the 450 GeV of the SPS to the 6.5 TeV of the LHC by a system of eight radio frequency (RF) cavities. These RF cavities oscillate at 400 MHz at a maximum amplitude of 2 MV, for a total of 16 MV per beam, increasing the proton energy by about 0.5 MeV per revolution.

The phase of the RF waveform is carefully modulated to create and maintain bunches of protons, and to accelerate them to and maintain them at the desired energy. The 3564 bunch places of the LHC are separated by 25 ns, meaning a bunch place passes through a particular RF cavity at a frequency of 40 MHz. The RF cavity frequency is ten times that: 400 MHz. This divides a bunch place into ten RF buckets, which are individual regions of space in which a bunch of protons can be confined, the shape and size of which are determined by the RF

## The CERN accelerator complex *Complexe des accélérateurs du CERN*

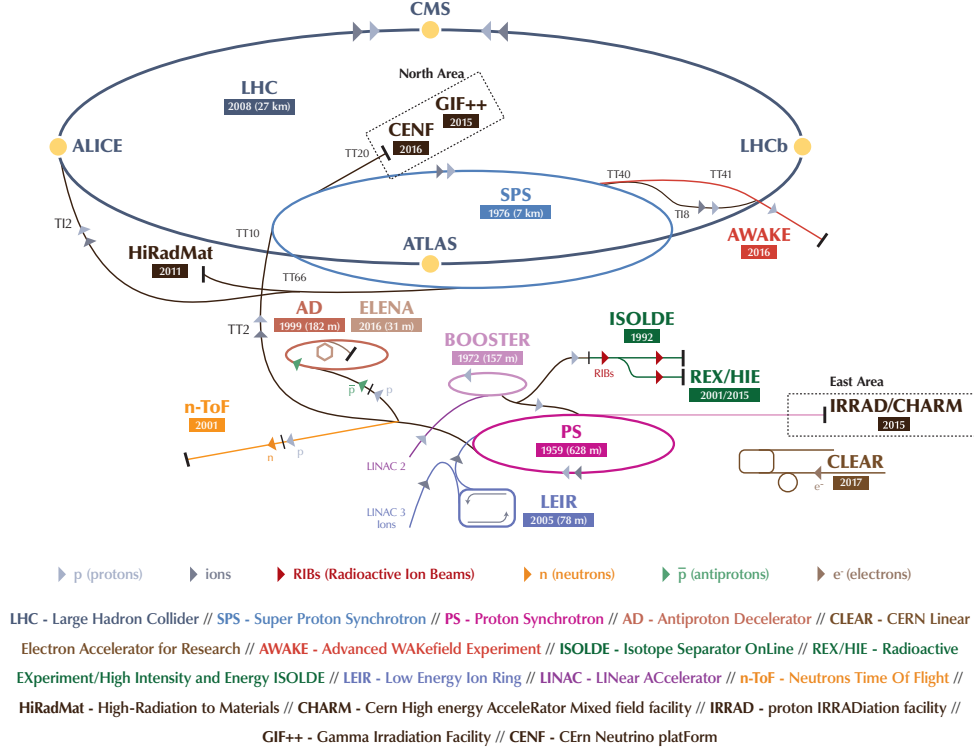


Figure 3.1: Overview of the CERN accelerator complex, reproduced from Reference [17].



Figure 3.2: Summary of the injection chain of the LHC, including the proton beam energy, which increases at each step, and (for the PS and SPS) the synchrotron cycle time. Multiple cycles of the PS and SPS are required to fill the LHC.

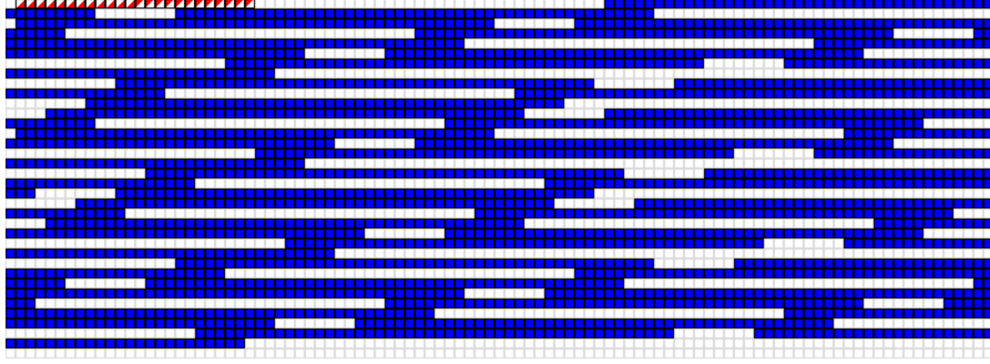


Figure 3.3: Diagram of the bunch places of an LHC proton beam, showing filled bunches (blue) and empty places (gray). The grid is  $99 \times 36 = 3564$  squares, i.e. they wrap around. Reproduced from Reference [18] for Fill 5423 of the LHC in October 2016.

voltage amplitude and the number of bunch places. When coasting (at collision energy), a (hypothetical) particle with energy such that the RF frequency is exactly an integer multiple of the orbit frequency, synchronized such that the particle passes through the RF cavity at exactly a time of zero voltage within the RF waveform is called a synchronous particle. Particles with higher or lower energy or arriving early or late with respect to the synchronous particle will experience a non-zero voltage and therefore experience a restoring force. These particles oscillate longitudinally around the synchronous particle. The size of an RF bucket is given as an area in energy-time phase space, defined with respect to a synchronous particle: it is parametrized by the maximum energy deviation of a particle within the bunch from that of the synchronous particle (in eV) and the maximum deviation in arrival time of a particle within the bunch from that of the synchronous particle (in seconds). A proton bunch occupies a single RF bucket and is depicted by an ellipse within the bucket in energy-time phase space, representing the trajectory of particles within the bunch. An LHC proton bunch has a  $4\sigma$  bunch length of 1.06 ns at collision energy. Figure 3.4 depicts an RF bucket, football-shaped in energy-time phase space, as well as the ellipse contained within it that represents the proton bunch. The area of the bunch in this diagram is called the longitudinal emittance, with units of  $\text{eV} \cdot \text{s}$ .

Beams of particles in the LHC are steered by a network of 1232 superconducting dipole

magnets, interspersed with quadrupole magnets for focusing. The superconducting wire windings in these magnets are made of niobium-titanium, cooled by liquid helium to their operating temperature of 1.9 K, and producing a magnetic field of 8.33 T. Figure 3.5 is a diagram of the cross section of an LHC dipole magnet as well as a visualization of the magnetic field lines within the beam pipe.

### 3.1.3 Instantaneous and Integrated Luminosity

The number of events per second generated by the LHC for a process of cross section  $\sigma$  is

$$\frac{dN}{dt} = \mathcal{L}\sigma \quad (3.1)$$

where  $\mathcal{L}$  is the instantaneous luminosity, and depends only on machine parameters:

$$\mathcal{L} = \frac{N_p^2 N_b f \gamma}{4\pi\epsilon\beta^*} R \quad (3.2)$$

where  $R$  is a geometrical factor accounting for the beam crossing angle,

$$\frac{1}{R} = \sqrt{1 + \left(\frac{\theta\sigma_z}{2\sigma^*}\right)^2} \quad (3.3)$$

The design LHC beam parameters in Equations 3.2–3.3 are defined and summarized in Table 3.1 [19, 22]. This yields a peak instantaneous luminosity of  $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ .

Proton collisions lead to a natural decrease in luminosity over time, with a luminosity lifetime of 15–25 hours. The integral of the instantaneous luminosity over time is called the integrated luminosity,  $\mathcal{L}_{\text{int}}$ . The total number of events generated by the LHC for a process of cross section  $\sigma$  is then given by the integrated luminosity times the cross section, so that

$$N = \sigma \int \mathcal{L} dt = \sigma \mathcal{L}_{\text{int}} \quad (3.4)$$

Since the pp interaction cross section is a constant, the total integrated luminosity delivered to (and recorded by) an experiment is therefore also a measure of the number of events and hence the amount of data recorded by the experiment. For convenience with handling large exponents, the integrated luminosity is given in inverse femtobarns:  $1 \text{ fb} = 10^{-39} \text{ cm}^2$ . Figure 3.6 is a plot of the total pp integrated luminosity over time, by year and calendar

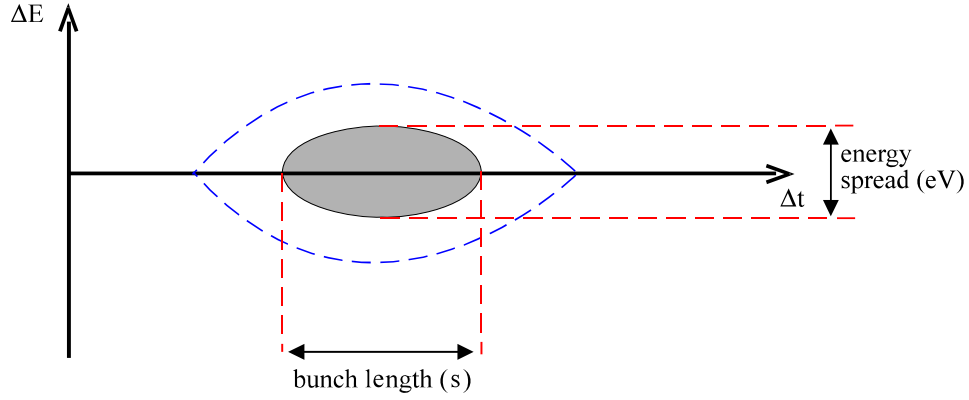


Figure 3.4: Phase space diagram of an RF bucket and the elliptical boundary of a proton bunch within the bucket, reproduced from Reference [19] with minor typographical corrections.

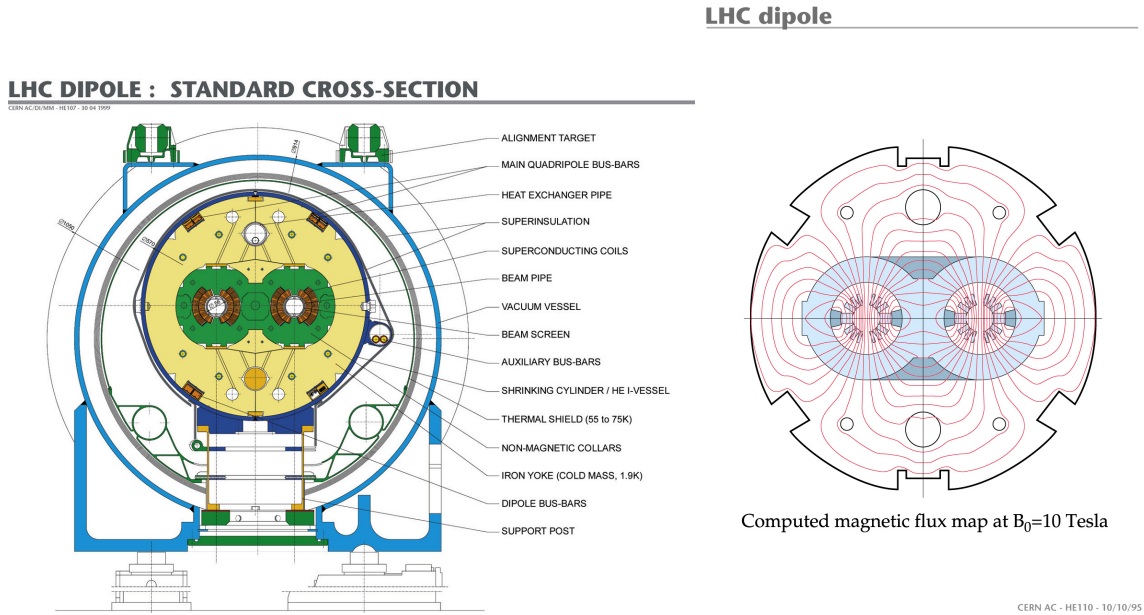


Figure 3.5: (left) Diagram of cross section of an LHC steering dipole magnet, reproduced from Reference [20], showing the two beam pipes and the windings. (right) Depiction of the magnetic field lines in the magnets, reproduced from Reference [21], showing them to point in the vertical direction within the beam pipe, as is required to circulate particles along the surface of the earth.

Symbol	Name	Value
$N_p$	protons per bunch	$1.15 \times 10^{11}$
$N_b$	bunches per beam	2808
$f$	revolution frequency (1/24.95 ns/3564)	11.25 kHz
$\gamma$	relativistic Lorentz factor for protons ( $E_p/m_p$ )	7461
$\epsilon$	normalized transverse beam emittance	$3.75 \mu\text{m}$
$\beta^*$	optical $\beta$ function (amplitude of betatron oscillations)	55 cm
$\theta$	beam crossing angle	$285 \mu\text{rad}$
$\sigma_z$	longitudinal RMS bunch length	7.55 cm
$\sigma^*$	transverse RMS beam size	$16.7 \mu\text{m}$

Table 3.1: LHC nominal design beam parameters, reproduced from Reference [22].

month, for the entirety of Run 1 and Run 2. This thesis presents two analyses using data taken by the CMS experiment in 2016: a search for displaced dimuons using the full 2016 integrated luminosity of  $35.9 \text{ fb}^{-1}$ , and a study of neutron-induced background in muon chambers using one era of 2016 data taking, corresponding to  $8.73 \text{ fb}^{-1}$ .

## 3.2 The CMS Detector

### 3.2.1 Introduction

CMS is located at Point 5 of the LHC in the commune of Cessy in eastern France, at the level of the LHC beam line, 100 m underground. Its overall shape is a cylinder 15 m in diameter and 21.6 m in length, weighing 14,000 tons. It is named for three of its distinguishing features: its **compactness**, its **muon** system, and its **solenoid** magnet. A large magnetic field with high bending power is required to precisely measure the momentum of high-energy charged particles. This informs a choice of superconducting technology, and so a superconducting solenoid magnet sits at the heart of the cylindrically symmetric CMS detector, producing a continuous magnetic field of 3.8 T. CMS is quite compact for all the detector material it

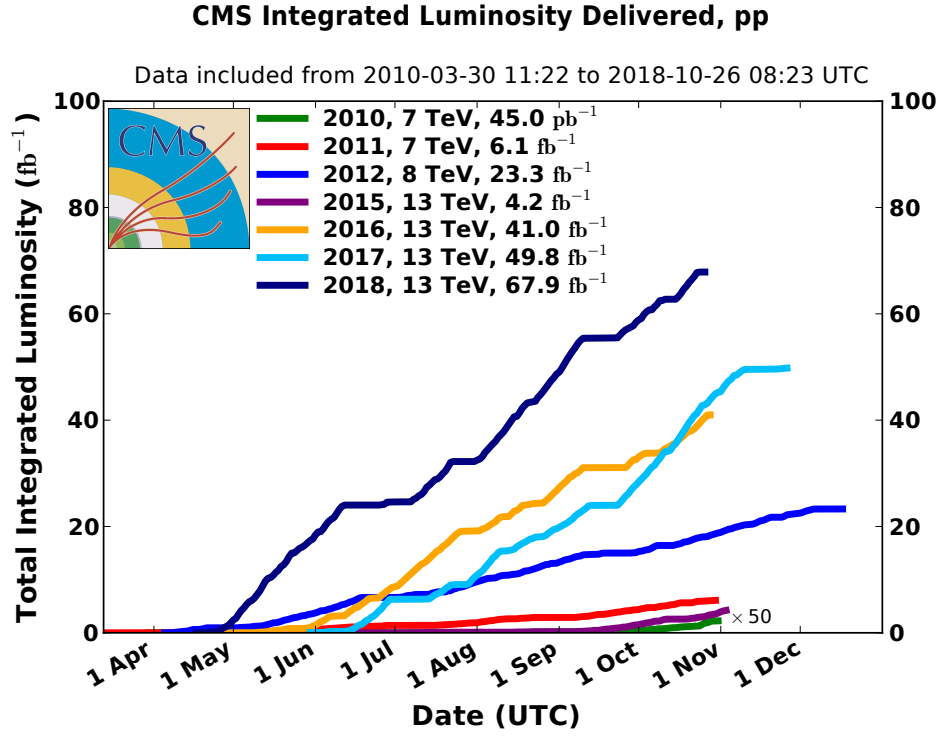


Figure 3.6: Total pp integrated luminosity recorded by CMS during Run 1 and Run 2 by year and month, reproduced from Reference [23].



contains, especially compared to ATLAS. Notably, the bulk of the CMS hadronic calorimeter is completely contained the solenoid magnet, with important consequences for its design. The muon system, on the other hand, is outside the solenoid. Muons are one of the five general categories of particles directly detected by CMS, and unique among them in that they typically neither stop nor decay within the boundaries of the detector. They are less subject to energy losses when passing through detector material than electrons and so provide a powerful lens with which to study high-energy processes in the presence of high background. The outermost bulk of CMS is thus a dedicated system for identifying and measuring muons, consisting of three kinds of gas ionization detectors [24].

The CMS detector is structured like an onion, in layers, consisting of the following basic subsystems, ordered from innermost to outermost:

- silicon tracker
- electromagnetic calorimeter
- hadronic calorimeter
- superconducting solenoid magnet
- muon system

Figure 3.7 is a diagram of a slice of the CMS detector illustrating these layers. It also shows the detector signatures of the general categories of particles directly detected by CMS:

- electrons
- photons
- charged hadrons
- neutral hadrons
- muons

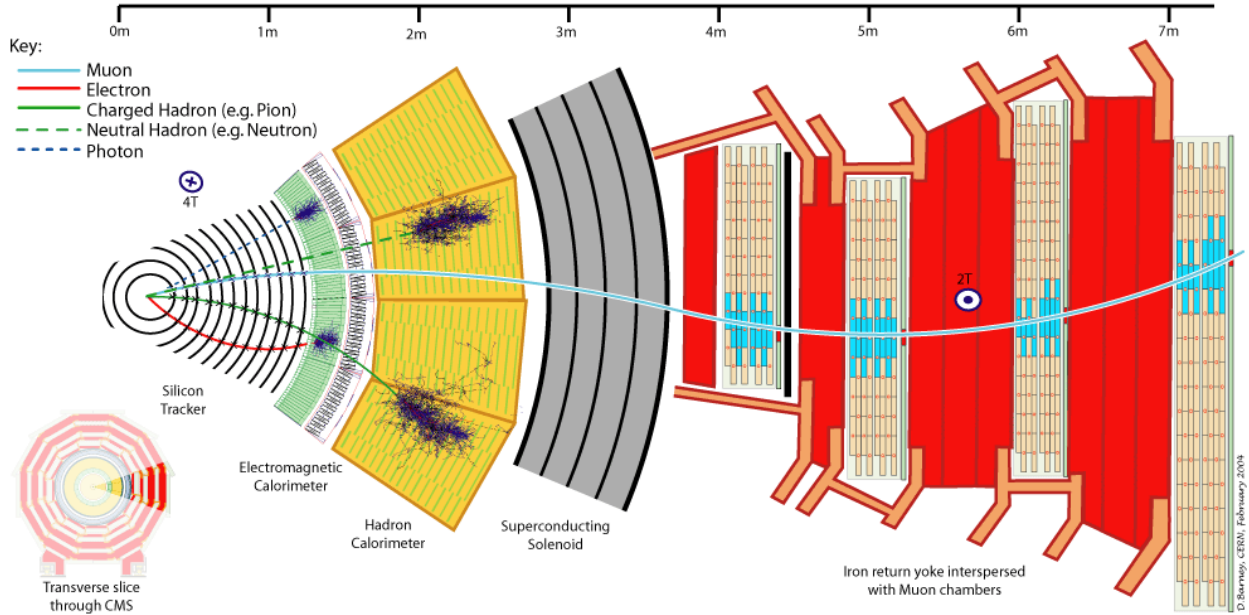


Figure 3.7: Transverse slice of the CMS detector, reproduced from Reference [25], illustrating the basic structure of CMS and the shapes of the tracks and energy deposits formed by the five general categories of particles directly detected by CMS by one or more of its subsystems.

The charged particles—electrons, muons, and charged hadrons—are detected as they form curved, helical tracks in the silicon tracker (and in the case of muons, in the muon system as well). Electrons, charged hadrons, and the neutral particles—photons and neutral hadrons—leave energy deposits in the electromagnetic and hadronic calorimeters.

### 3.2.2 Coordinate System

The origin of the coordinate system used by CMS is the nominal pp collision point. The  $y$ -axis points upwards, the  $x$ -axis points radially inwards towards the center of the LHC, approximately south, and thus the  $z$ -axis points approximately west. The azimuthal angle  $\phi$  about the  $z$ -axis is measured from the  $x$ -axis and the radial coordinate in the  $xy$ -plane is denoted  $r$ . The polar angle measured from the  $z$ -axis is denoted  $\theta$ . However, a more conventional coordinate used in hadron collider physics is the pseudorapidity  $\eta$ , defined as

$$\eta = -\ln \tan \left( \frac{\theta}{2} \right) \quad (3.5)$$

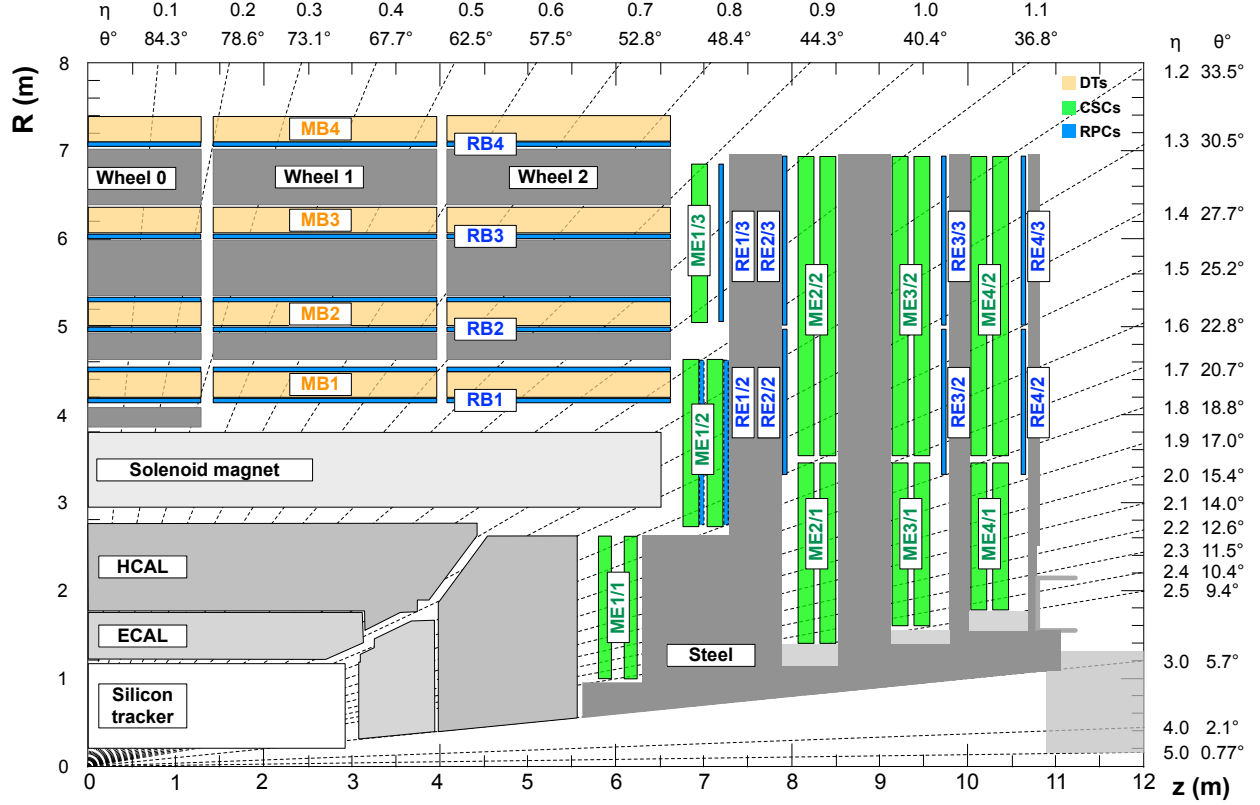


Figure 3.8: Schematic diagram of one quadrant of CMS in  $r$ - $z$ , reproduced from Reference [27], illustrating the position of all the subsystems with respect to the coordinate system, in both the barrel and the endcap.

For a particle of three-momentum  $\mathbf{p}$  with  $z$ -component  $p_z$ , pseudorapidity can be written

$$\eta = \frac{1}{2} \ln \left( \frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z} \right) \quad (3.6)$$

This form elucidates its relationship to rapidity (along the  $z$ -axis),

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right), \quad (3.7)$$

which is a quantity Lorentz invariant under boosts in the  $z$ -direction [26]. The pseudorapidity has the advantage that it converges to rapidity (along the  $z$ -axis) in the high-velocity, low-mass limit (as  $|\mathbf{p}| \rightarrow E$ ), and is only dependent on the polar angle  $\theta$  and not on the energy of the particle. Figure 3.8 is a schematic diagram of one quadrant of CMS, showing the placement of the components of CMS within the coordinate system.

The diagram illustrates that  $\eta = 0$  points upwards, while  $\eta \rightarrow \infty$  points along the  $z$ -axis. For this reason the “forward” regions close to the beam line (which are also less instrumented) are referred to as “high  $\eta$ .” The diagram also illustrates two distinct loci within the CMS detector: the barrel, which covers approximately (depending on subsystem) the region  $|\eta| < 1.2$ , and the endcap, which covers approximately (depending on subsystem) the region  $1.2 < |\eta| < 3$ .

### 3.2.3 Silicon Tracker

The inner tracking system of CMS must provide precise measurements of charged particle trajectories and reconstructions of secondary vertices of (on average) a thousand particles every 25 ns bunch crossing, exposed to the full flux of the radiation of the LHC. These requirements on precision, speed, and radiation hardness inform a choice of silicon detector technology. The ionization induced by an energetic charged particle passing through produces electron-hole pairs, which are measured as current in the presence of an applied voltage.

The CMS tracker consists of two silicon detector technologies: pixels and strips. The innermost component of the tracker, closest to the collision point, is the pixel detector, consisting of three barrel layers and two endcap disks on each side of the barrel. The 66 million silicon pixel sensors are  $n$ -on- $n$  devices, measuring  $100\,\mu\text{m} \times 150\,\mu\text{m}$ , giving a spatial resolution of 15–20  $\mu\text{m}$ . Just outside the pixel detector is the strip detector, consisting of ten barrel layers and twelve endcap disks on each side of the barrel. The 9.6 million silicon strip sensors are  $p$ -on- $n$  type microstrip sensors, manufactured on 6 in wafers, varying in width from 80–180  $\mu\text{m}$ . The CMS tracker measures transverse momentum ( $p_T$ ) to a resolution of 1% for 100 GeV particles. Containing altogether 205 m<sup>2</sup> of silicon sensors with 75 million individual channels, the CMS tracker is the largest silicon detector in the world [24, 28, 29].

### 3.2.4 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) was designed to achieve the excellent energy resolution critical for observing the decay of the SM Higgs boson to two photons:  $H \rightarrow \gamma\gamma$ . The ECAL is a homogeneous calorimeter consisting of about 76,000 scintillating lead tungstate ( $\text{PbWO}_4$ ) crystals, a material chosen for its high density and short radiation length. Electrons or photons passing through the detector material result in a cascade of electromagnetic interactions, producing a shower of particles culminating in a release of energy proportional to the energy of the incident particle. The corresponding photons are then measured by photodetectors installed on each crystal [24, 30, 31].

### 3.2.5 Hadronic Calorimeter

The hadronic calorimeter (HCAL) is a sampling calorimeter: it consists of repeating, alternating layers of an absorber, which interacts with an incident particle to produce more particles of lower energy, and an active medium, which provides a detectable signal. This is in contrast to a homogeneous calorimeter, like the ECAL, in which a single type of material performs both functions.

The HCAL consists of four parts: the barrel and endcap HCAL (HB and HE), the outer HCAL (HO), and the forward HCAL (HF). In the HB and HE, the absorber material consists of thick tiles of brass, and the active medium consists of thinner tiles of scintillating plastic with wavelength-shifting readout fibers. Brass is a dense material with many nuclei to interact strongly with incident hadron showers. It is also non-magnetic, a necessary property of the absorber for the HB and HE, which lie within the solenoid and experience its full 3.8 T magnetic field.

As the HB and HE lie within the solenoid and hence are only about 6 interaction lengths thick, the first layer of the muon system is instrumented with scintillator tiles, treating the solenoid as an additional absorber. This “tail catcher” is known as the outer calorimeter, or HO.

The HF is located 11 m from the interaction point, in the pseudorapidity range of approximately  $3.0 < |\eta| < 5.0$ . This forward region experiences very high levels of LHC radiation and thus is constructed out of radiation-hard materials: steel for the absorber and quartz fiber for the active material, which detects the Cherenkov radiation produced by energetic jets [24, 32, 33].

### 3.2.6 Solenoid Magnet

The CMS magnet is a superconducting solenoid, 12.5 m long with an inner diameter of 6.3 m. Liquid helium cools the magnet to its superconducting operating temperature of 4 K. It draws 19 kA of current and is the largest magnet in the world in terms of its 2.6 GJ of stored energy.

A powerful magnetic field is crucial for precise momentum measurement of high-energy charged particles. As an introduction to issues related to muon  $p_T$  measurement, consider a charged particle of charge  $q$  and transverse momentum  $p_T = \mathbf{p} \cdot \hat{\mathbf{r}}$  in a uniform magnetic field of strength  $B$  pointing in the  $z$ -direction, which is a good model for a charged particle in the CMS tracker. Such a particle travels in a helix of (signed) radius

$$R = \frac{p_T}{qB} \quad (3.8)$$

The radius of curvature of a charged particle in a magnetic field is thus proportional to its transverse momentum. However, what is measured directly is not the radius of curvature, but rather the positions of the hits, whose resolution functions are Gaussian. To illustrate an important consequence of this, consider Figure 3.9, which depicts the circular arc left by a charged particle in the tracker. Let  $L$  be the length of the chord joining the outermost points, which is also the radius of the tracker. Let  $s$  be the sagitta, i.e. the distance between the midpoint of the chord and the center of the arc. Then  $R$ , the radius of curvature, can be computed from  $L$  and  $s$  as follows:

$$R = \frac{L^2}{8s} + \frac{s}{2} \approx \frac{L^2}{8s} \quad (3.9)$$

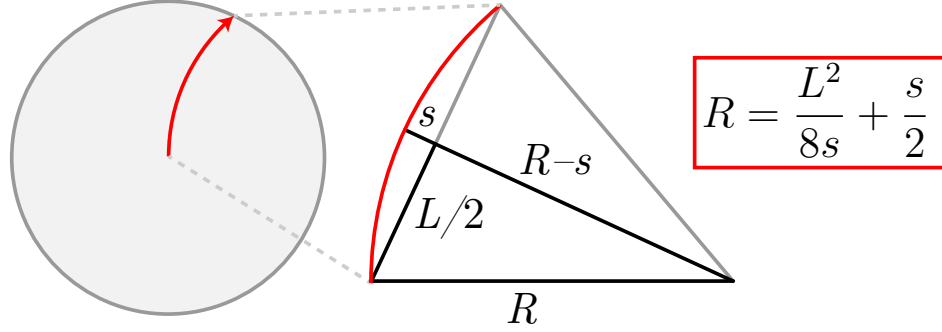


Figure 3.9: Radius of curvature from sagitta. On the left is a depiction of a charged particle leaving a circular arc (red) in the CMS tracker (gray). The length of the chord is  $L$ , which is also typically the radius of the tracker. The sagitta  $s$  is approximately inversely proportional to the radius of curvature  $R$ , with important consequences for the uncertainty in the desired  $p_T$  measurement.

Since  $s$  is typically small compared to  $L$ , the  $s/2$  term can be dropped. Then  $q/p_T$  is approximately proportional to  $s$ :

$$\frac{q}{p_T} = \frac{1}{BR} \approx \frac{8s}{BL^2} \quad (3.10)$$

Since  $s$  is linear in the position measurements, its distribution is also Gaussian, and therefore it is the distribution of  $q/p_T$ , and not  $p_T$ , that is Gaussian. The uncertainty on a measurement of  $p_T$  obtained from applying standard error propagation to  $q/p_T$  must therefore be considered carefully, as it does not describe standard deviations of a variable distributed as a Gaussian.

### 3.2.7 Muon System

As mentioned in Section 3.2.1, muon detection is a powerful tool for studying high-energy processes in the presence of high background. Because of the amount of material in the inner subsystems, typically mostly muons travel through the solenoid. A track in the muon system therefore is associated with a muon. Hadronic “punchthrough” in the muon system is minimal.

The muon system has three tasks: triggering (Section 3.2.8), muon identification, and muon reconstruction. As with the other subsystems, the shape of the solenoid informs a design of a cylindrical barrel section and an endcap disk section. Both sections consist of four stations of muon detectors, concentric for the barrel and sequential for the endcap. The muon system consists of three kinds of gaseous ionization detectors: drift tubes (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC).

The chambers of the muon system are embedded within a set of steel disks in the endcap and concentric twelve-sided steel cylinders in the barrel, referred to as the return yoke. The return magnetic flux of the solenoid is mostly contained within this return yoke [24, 34].

### 3.2.7.1 Drift Tubes

The barrel region of the muon system consists of four stations instrumented with 250 drift tube chambers. Drift tubes were chosen to be the tracking detectors in the barrel region in light of the low expected rate and relatively low intensity of the local magnetic field. Figure 3.10 is a diagram showing the principle of operation of a drift tube cell. A cathode tube with cross-sectional dimensions  $42\text{ mm} \times 13\text{ mm}$  contains an anode wire (operating at 3600 V) under tension. The tube is filled with a gas mixture of 85% Ar and 15% CO<sub>2</sub>. An energetic muon passing through this cell ionizes the gas, and the resulting electrons drift towards the wire. Measuring the drift time (a maximum of 380 ns) yields a measurement of position within the cell.

The smallest independent unit of a DT is a superlayer (SL), consisting of four layers of drift cells staggered by a half cell. A drift tube chamber consists of three (or two) SLs. The wires in the two outer SLs are parallel to the beam line, and provide a track measurement in the  $r$ - $\phi$  (bending) plane. The wires in the inner SL are orthogonal to the beam line, and measure the  $z$ -position along the beam. This inner SL is not present in the fourth muon station, which consequently only measures the  $\phi$  coordinate.



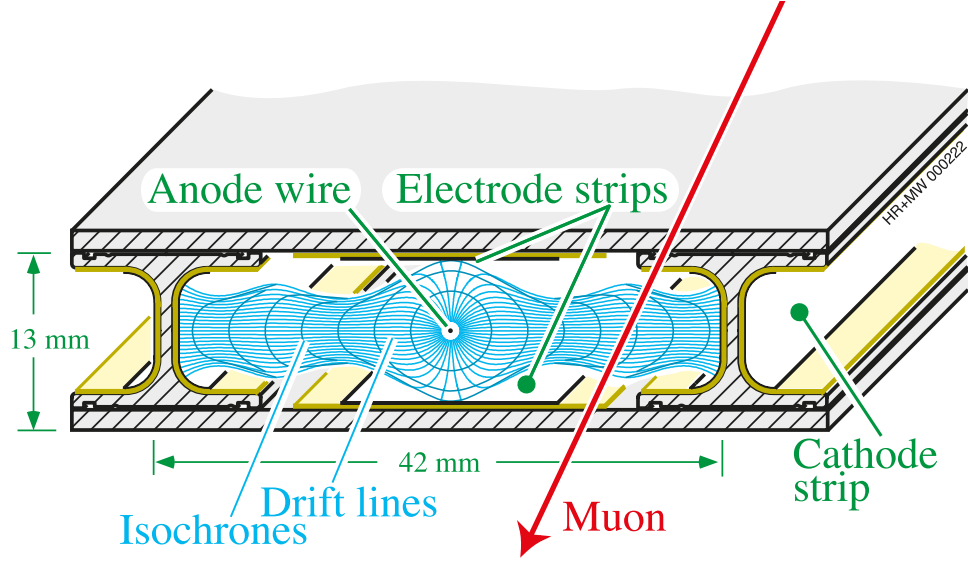


Figure 3.10: Principle of operation of a drift tube cell, reproduced from Reference [35], showing the structure of a cell, as well as the drift lines and isochrones.

### 3.2.7.2 Cathode Strip Chambers

The endcap region of the muon system consists of four stations instrumented with 540 cathode strip chambers. Cathode strip chambers were chosen to be the tracking detectors in the endcap region for their excellent position resolution in the  $\phi$  direction achieved by precision cathode charge readout and interpolation. The CSCs are arranged in circular disks. Each CSC consists of six layers, each layer lying in an  $r$ - $\phi$  plane of CMS, consisting of a gas mixture of 50% CO<sub>2</sub>, 40% Ar, and 10% CF<sub>4</sub> in between a plane of copper cathode strips and a plane of anode wires, operating at 2900–3600 V. An energetic muon passing through a CSC ionizes the gas, and the resulting electrons drift towards the wires, causing an avalanche of charge that induces an opposite charge on the cathode strips. Interpolating these charges yields a precise localization of the avalanche. A more detailed overview of CSCs in the context of a study of neutron-induced background is given in Section A.2.

### 3.2.7.3 Resistive Plate Chambers

Interspersed throughout both the barrel and endcap muon system are 480 and 576 resistive plate chambers, respectively, whose purpose is a fast time response and resolution comparable to that of scintillators. RPCs therefore are part of a dedicated muon trigger for identifying muon tracks and assigning the bunch crossing with high efficiency. An RPC consists of two parallel plates of phenolic resin coated with conductive graphite, with a 2 mm gap filled with a gas mixture of 95.2% freon ( $\text{C}_2\text{H}_2\text{F}_4$ , known as R134a), 3.5% isobutane ( $\text{i-C}_4\text{H}_{10}$ ), and 0.3% sulfur hexafluoride ( $\text{SF}_6$ ). An energetic muon crossing the chamber ionizes the gas and induces an image charge, which is sampled and read out. The RPCs have a time resolution of about 2 ns, a much shorter time than the 25 ns between LHC bunch crossings, but much coarser position resolution than DTs or CSCs.

### 3.2.7.4 Muon Momentum Resolution

An important consequence of a dedicated muon system for reconstructing muons is that it improves the  $p_{\text{T}}$  resolution for muons at high muon  $p_{\text{T}}$ . Figure 3.11 shows plots of the muon  $p_{\text{T}}$  resolution using the inner tracking system only, using the muon system only, and combining both, as a function of  $p_{\text{T}}$ . At lower  $p_{\text{T}}$  values, the tracker is sufficient to precisely measure the curvature of the muon track, and the tracker-only resolution dominates. But at higher  $p_{\text{T}}$  values, the tracker track is much straighter, and the additional information about muon position obtained from the muon system results in a  $p_{\text{T}}$  resolution superior to that obtained by using either system alone.

## 3.2.8 Trigger System

The LHC provides pp collisions every 25 ns, corresponding to a crossing frequency of 40 MHz. It is impossible to process and store this large amount of data synchronously with such a high rate, and so a rate reduction must be achieved, selecting a subset of the most interesting candidate events for physics analysis. This rate reduction is performed by the CMS trigger system, an elaborate system of hardware and software that analyzes every bunch crossing

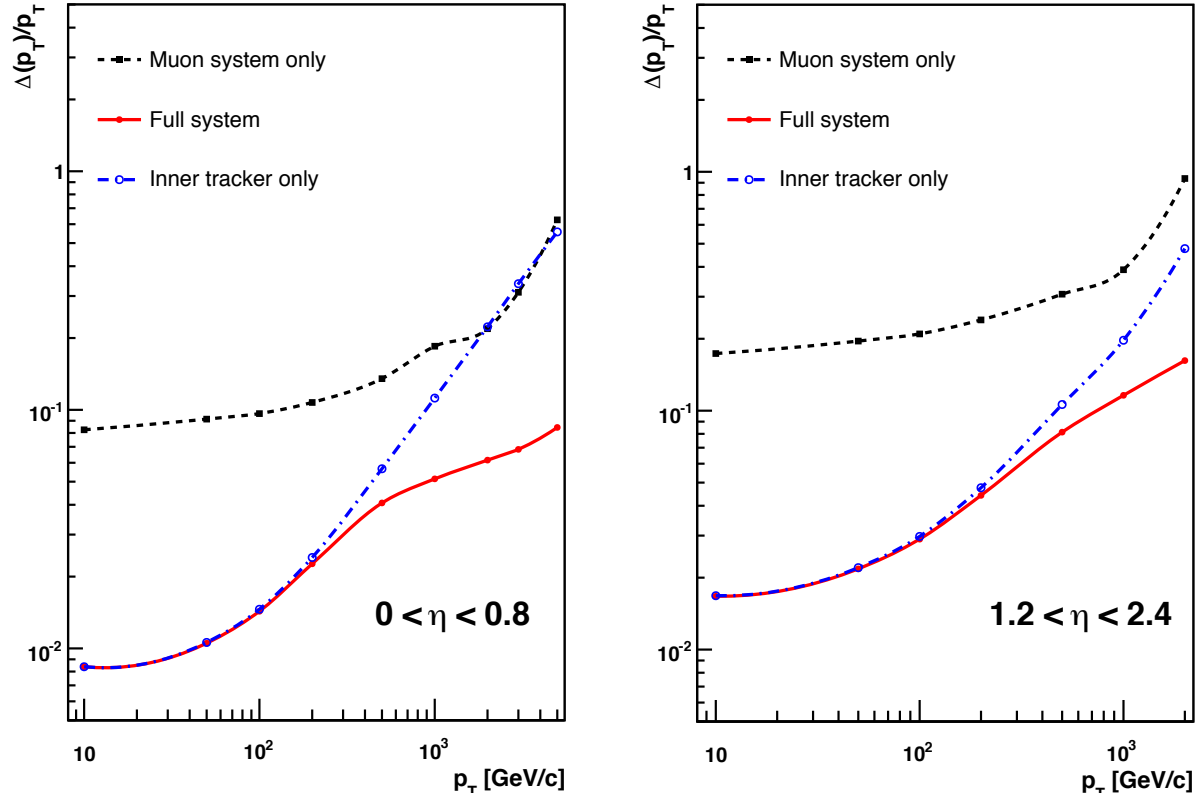


Figure 3.11: Muon  $p_T$  resolution as a function of muon  $p_T$ , for (left) the barrel and (right) the endcap, using the inner tracking system only, using the muon system only, and combining both. Reproduced from Reference [24].

and “triggers” on potentially interesting events. The CMS trigger system operates in two steps. The Level-1 (L1) trigger makes its decisions at the hardware level with custom-built programmable electronics, synchronously with the LHC, and is designed to reduce the rate from 40 MHz to 100 kHz. The High-Level Trigger (HLT) makes its decisions at the software level, performing its computations in real time with respect to the full rate of the LHC, using a processor farm located on the surface at Point 5. HLT processing operates more slowly than the L1 trigger, but faster than the full event readout, serving to reduce the rate from 100 kHz to the data acquisition rate of hundreds of Hz [24, 36].

The L1 trigger consists of a muon trigger and a calorimeter trigger, organized into local, regional, and global components. Both the ECAL and all components of the HCAL participate in the calorimeter trigger, and all three muon chambers—DTs, CSCs, and RPCs—participate in the muon trigger, the first muon trigger to measure momentum at the hardware Level-1. The local components are the lowest level, based on energy deposits in the calorimeters and hits and segments in the muon system. Regional components combine the information from the local components using pattern recognition and track finding and rank trigger objects in small regions. The global components determine the highest-rank objects and transfer them to the global trigger of the L1 system. The global trigger then makes the decision to reject or accept the event at Level-1. A Level-1 Accept (L1A) decision is communicated to the subsystems and the event is passed to the HLT for further evaluation. The L1 trigger must analyze every bunch crossing, doing so with a maximum latency of  $3.2\,\mu\text{s}$ . Figure 3.12 depicts the architecture of the L1 trigger.

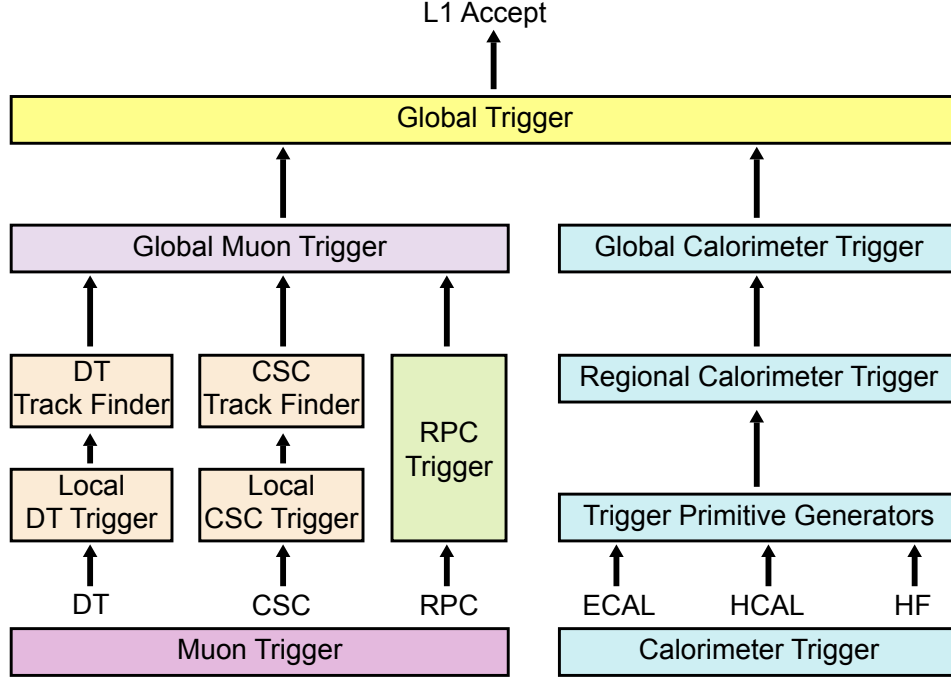


Figure 3.12: Architecture of the L1 trigger system, based on a figure from Reference [24], depicting the local, regional, and global components of the muon and calorimeter triggers, the information from which is processed by the global L1 trigger for trigger decision at L1.

The L1 and HLT menus consists of paths that define criteria on which to trigger. These criteria are the first step to any physics analysis, and so can include requirements on  $p_T$ , on numbers of muons or jets, on the geometry and angles between them, and much more. The HLT uses information from the entire detector, with high-level algorithms that make a more precise decision than the coarser L1 trigger. Each HLT path is seeded by one or more L1 paths and is designed to be inclusive and general while keeping the event rate in the data acquisition system within its maximum rate.

## CHAPTER 4

# Search for Long-Lived Particles Decaying to Displaced Dimuons in the CMS Muon System at $\sqrt{s} = 13$ TeV

*When you have eliminated the impossible, whatever remains, however improbable, must be the truth.*

—Sherlock Holmes, *The Sign of the Four*, by Sir Arthur Conan Doyle

This chapter presents a search for long-lived particles decaying to two muons in the CMS detector with data taken in 2016 at center-of-mass energy  $\sqrt{s} = 13$  TeV corresponding to  $36.3\text{ fb}^{-1}$  of integrated luminosity during Run 2 of the LHC. The reconstructed muons form dimuon vertices displaced with respect to the pp collision interaction point, giving rise to a variety of interesting experimental challenges. This analysis focuses on muons reconstructed using only the muon system of CMS, and is a continuation and extension of two CMS analyses [1, 2] performed with data taken at center-of-mass energy  $\sqrt{s} = 8$  TeV during Run 1 of the LHC. The results are interpreted in the context of the BSM Higgs model introduced in Chapter 2, but the search is meant to be inclusive and model-independent.

## 4.1 Trigger, Data, and Simulation

### 4.1.1 Trigger

The choices of trigger paths determine which events in data are used in the analysis. Typical CMS analyses using muons search for new particles produced in the hard interactions of pp collisions that decay quickly into muons. Such muons appear to originate directly from the point of interaction of the crossing beams, or beam spot. For this reason, the reconstruction

of muon tracks used by most trigger paths includes knowledge of the beam spot, referred to as a beam spot constraint. But muons produced from the decay of a long-lived particle may not point towards the beam spot. Therefore, dedicated triggers without beam spot constraints are necessary to collect data containing long-lived particles.

The HLT path that collected the 2016 data studied in this analysis was

`HLT_L2DoubleMu28_NoVertex_2Cha_Angle2p5_Mass10`

The abbreviations in this expression define the various criteria:

- **L2DoubleMu28:** requirement that two muons be reconstructed using only the muon system, each with a  $p_T$  of at least 28 GeV. The muon system requirement is the first step towards ensuring that the triggered objects are really muons, as well as allowing the analysis to be sensitive to a range of possible lifetimes. The  $p_T$  requirement keeps the rate of triggered events in the data acquisition system within its maximum rate and allows the analysis to focus on events in a regime with reduced contributions from complex QCD processes.
- **NoVertex:** requirement that no beam spot constraint be imposed in the muon reconstruction.
- **2Cha:** requirement that muon segments be present in at least two CSC or DT stations. This requirement selects muons with hits over multiple chambers, which is necessary to reconstruct them with high quality.
- **Angle2p5:** requirement that the 3D angle between the muons  $\alpha$  be at least 2.5 rad (equivalent to  $\cos \alpha > -0.8$ ). This requirement is intended to prevent cosmic muons from passing the trigger, as the reconstruction algorithms usually reconstruct cosmic muons as a pair of back-to-back muons, i.e.  $\alpha \approx \pi$  or  $\cos \alpha \approx -1$ .
- **Mass10:** requirement that the invariant mass of the two muon system  $m_{\mu\mu}$  be at least 10 GeV. As with the  $p_T$  requirement above, this requirement also serves to keep the event rate within acceptable thresholds and to reduce consideration of events dominated by complex QCD processes.

Dataset	Run Range	$\mathcal{L}_{\text{int}}$
DoubleMuon/Run2016B-07Aug17-ver2-v1/AOD	273150–275376	5.81 fb <sup>-1</sup>
DoubleMuon/Run2016C-07Aug17-v1/AOD	275656–276283	2.62 fb <sup>-1</sup>
DoubleMuon/Run2016D-07Aug17-v1/AOD	276315–276811	4.28 fb <sup>-1</sup>
DoubleMuon/Run2016E-07Aug17-v1/AOD	276831–277420	4.03 fb <sup>-1</sup>
DoubleMuon/Run2016F-07Aug17-v1/AOD	277932–278808	3.12 fb <sup>-1</sup>
DoubleMuon/Run2016G-07Aug17-v1/AOD	278820–280385	7.66 fb <sup>-1</sup>
DoubleMuon/Run2016H-07Aug17-v1/AOD	281613–284044	8.77 fb <sup>-1</sup>
<b>Total</b>		<b>36.30 fb<sup>-1</sup></b>

Table 4.1: CMS datasets used by the analysis, including dataset names, the “Muon Physics” run ranges as certified by CMS for physics analysis, and integrated luminosities, taken from Reference [37] and following embedded links. Integrated luminosity values were obtained from running the standard `brilcalc` prescription found at Reference [38].

#### 4.1.2 Data Samples

The analysis uses data taken from pp collisions at the LHC center-of-mass energy of 13 TeV during the data taking period of 2016, corresponding to an integrated luminosity of 36.3 fb<sup>-1</sup>. The CMS datasets used are reprocessed under the CMS software version `CMSSW_8_0_31` in August 2017, known as the “re-reco” data. CMS data are organized into runs consisting of many individual lumi sections, which are time intervals of data taking at CMS corresponding to approximately 23 s. A dedicated data quality management team certifies CMS data by producing lists of run numbers approved for physics analysis. Table 4.1 lists the names of the datasets used in this analysis, their certified run range, and their corresponding integrated luminosity for each of the data taking eras of 2016.

#### 4.1.3 Monte Carlo Simulation Samples of Signal and Background

Comparing theoretical predictions to CMS data is a complex task that must connect scattering amplitudes of the hard process underlying the pp collision computed via perturbative



methods in quantum field theory, to the production of particles and their passage through the material of CMS, to the response and subsequent data obtained by the CMS detectors. Numerical simulation is used to obtain these predictions. Because Monte Carlo methods are used to model stochastic effects at each stage, this simulation is described as Monte Carlo (MC) simulation. This analysis uses simulation of signal models as well as simulation of background processes in its studies.

Table 4.2 lists the MC simulation samples of the BSM Higgs  $H \rightarrow XX$  benchmark signal model described in Chapter 2. Both  $4\mu$  final states, in which both long-lived particles decay to two muons, and  $2\mu$  final states, in which one long-lived particle decays to two muons and the other does not, are simulated. Samples were produced for several combinations of BSM Higgs masses ( $m_H$ ), long-lived particle masses ( $m_X$ ), and  $c$  times the long-lived particle mean proper lifetimes ( $c\tau$ ). In order to study decays occurring everywhere from near the beam spot to the beginning of the muon system, the long-lived particle lifetimes were chosen such that the mean decay lengths in the transverse plane are 3 cm, 30 cm, and 250 cm in the laboratory frame. These signal samples can be reweighted to study additional intermediate lifetimes.

The main backgrounds for this analysis are

- the Drell-Yan process yielding dileptons ( $pp \rightarrow Z/\gamma^* \rightarrow \ell\ell$ ), especially from reconstruction mistakes;
- QCD processes yielding dileptons, especially from decays of bottom and charm quarks and hadron decays in flight; and
- cosmic muons reconstructed as dileptons.

Other backgrounds from top quark production and diboson production with leptonic decays produce negligible contributions. The CMS collaboration centrally produces MC simulation of background processes for physics analysis; this analysis uses the “Summer16” campaign produced for the winter conferences of 2017. Table 4.3 lists the MC simulation samples of background processes used in this analysis, with the total production cross section and the

$m_H$ (GeV)	$m_X$ (GeV)	$c\tau$ (mm)		
1000	350	35	350	3500
	150	10	100	1000
	50	4	40	400
	20	2	20	200
400	150	40	400	4000
	50	8	80	800
	20	4	40	400
200	50	20	200	2000
	20	7	70	7000
125	50	50	500	5000
	20	13	130	1300

Table 4.2: Simulated  $H \rightarrow XX$  signal samples used by the analysis, for several combinations of BSM Higgs mass ( $m_H$ ), long-lived particle mass ( $m_X$ ), and  $c\tau$ .

number of generated events, and an “equivalent luminosity” (defined in Section 4.1.4) that is important for understanding the statistical power of the simulation. These MC background samples are useful for many studies, but have some important limitations:

- The Drell-Yan and QCD processes are the dominant sources of background events in pp collisions and are simulated. However, the reconstruction mistakes producing this background may not be well modeled, and the simulation of QCD events is known to not accurately reproduce data in the tails of some distributions.
- The Drell-Yan and QCD samples also have limited statistical power, as the number of generated events is fewer than the number of expected events in 2016 data.
- The available MC simulation of cosmic muons does not include simulation of multiple cosmic muons from a single atmospheric shower initiated by a cosmic ray or of cosmic muons mixed with pp collisions.

Process	Kinematic Cuts	$\sigma$ (pb)	Events	$\mathcal{L}_{\text{int}}^{\text{eq}}$ (fb <sup>-1</sup> )
$Z/\gamma^* \rightarrow \ell\ell$	$10 \text{ GeV} < m_{\ell\ell} < 50 \text{ GeV}$	18810	30,935,823	1.20
$Z/\gamma^* \rightarrow \ell\ell$	$m_{\ell\ell} > 50 \text{ GeV}$	6225	122,547,040	13.19
QCD $\mu$ -enriched	$p_T > 20 \text{ GeV}$	302672.16	22,094,081	0.07
$t\bar{t} \rightarrow b\bar{b}WW \rightarrow b\bar{b}\ell\ell\nu\nu$		87.31	79,140,880	906.44
$tW, \bar{t}W$		35.8	13,885,924	387.87
WZ		47.13	6,995,142	84.82
ZZ		16.523	2,986,132	120.32
$WW \rightarrow \ell\ell\nu\nu$		12.178	1,999,000	164.15
W+jets		61526.7	29,804,825	0.48

Table 4.3: CMS simulation samples used by the analysis for each background process, along with the total production cross section and the number of generated events. The equivalent luminosity in 2016  $\mathcal{L}_{\text{int}}^{\text{eq}}$  is the proportionality factor that is used to scale simulated events to correspond to the total integrated luminosity in 2016 (36.3 fb<sup>-1</sup>).

For these reasons, MC simulation of background does not provide a good description of the expected background, so optimization of the analysis as well as evaluation of the expected background is consequently performed using data. MC simulation of background is therefore primarily used to study general trends and to gain insights.

Simulation of the hard process including matrix element computation is performed by POWHEG 2 [39, 40, 41] in all signal and background samples, except for the Drell-Yan ( $Z/\gamma^* \rightarrow \ell\ell$ ) sample, whose hard process is simulated by MADGRAPH5\_aMC@NLO 2.2.2 [42], and the W+jets sample, whose hard process is simulated by MADGRAPH 5 [43]. Simulation of parton showering and hadronization is performed by PYTHIA 8.212 using the CUETP8M1 tune [13, 14] in all signal and background samples except for the WW sample, whose parton showering is simulated by HERWIG++ 2.7.1 [44]. Simulation of the passage of particles through detector material and the induced response in detector electronics is performed by GEANT4 [45, 46, 47].

#### 4.1.4 Equivalent Luminosity and Event Weighting in MC Simulation

Because the number of simulated events is not equal to the number of expected events in data, comparing simulation to data requires that the contribution of each event be scaled by a weight factor depending on the cross section and the number of events. This weight factor can be understood in terms of an “equivalent integrated luminosity.” Given a number of generated events  $N_{\text{events}}$  and a cross section of  $\sigma$ , the equivalent luminosity is

$$\mathcal{L}_{\text{int}}^{\text{eq}} = \frac{N_{\text{events}}}{\sigma} \quad (4.1)$$

Then, when comparing to  $\mathcal{L}_{\text{int}}$  of data, the contribution of each simulated event is scaled by the weight factor

$$w = \frac{\mathcal{L}_{\text{int}}}{\mathcal{L}_{\text{int}}^{\text{eq}}} = \frac{\sigma \mathcal{L}_{\text{int}}}{N_{\text{events}}} \quad (4.2)$$

In practice, two additional subtleties modify Equations 4.1–4.2. First, corrections to the cross section  $\sigma$  when passing from leading-order to next-to-leading-order are given as a scaling factor  $k$ . This is accounted for by the substitution  $\sigma \rightarrow \sigma k$ . Second, matrix element computations performed by MADGRAPH5\_aMC@NLO assign a negative contribution to some events, according to the sign of the contributing diagrams. In such samples (in the case of this analysis, the Drell-Yan samples), the real number of events  $N_{\text{events}}$  is not the same as the number of generated events  $N_{\text{gen}}$ . If the number of events with positive weights and negative weights are  $N_+$  and  $N_-$ , respectively,

$$N_{\text{gen}} = N_+ + N_-, \quad N_{\text{events}} = N_+ - N_-$$

Then let  $f_{\text{neg}}$  be the fraction of generated events with a negative event weight, i.e.

$$N_- = f_{\text{neg}} \times N_{\text{gen}}$$

Then the real number of events

$$N_{\text{events}} = N_+ - N_- = N_{\text{gen}} (1 - f_{\text{neg}}) - f_{\text{neg}} N_{\text{gen}} = N_{\text{gen}} (1 - 2f_{\text{neg}}) \quad (4.3)$$

Any sample, with or without negative event weights, can thus be accounted for by the substitution  $N_{\text{events}} \rightarrow N_{\text{events}} (1 - 2f_{\text{neg}})$ . The expression for equivalent luminosity, Equation 4.1, is consequently modified, and the expression for the weight factor for simulated

events, Equation 4.2 is thus given by

$$w = \frac{\sigma k}{N_{\text{events}} (1 - 2f_{\text{neg}})} \times \mathcal{L}_{\text{int}} \quad (4.4)$$

## 4.2 Transverse Decay Length and Transverse Collinearity Angle

Particles produced from the hard interaction of pp collisions are usually associated with a vertex near the beam spot known as the primary vertex (PV). A long-lived particle produced from the PV will travel some distance in the detector before decaying into muons; this is observed as a displaced vertex. Consequently, a pair of reconstructed muon tracks are fit to a common vertex (CV), and the system of muons is referred to as a dimuon.

The transverse decay length,  $L_{xy}$ , is the distance between the PV and the CV in the transverse plane. Decays occurring at different  $L_{xy}$  can have significantly different properties. For example, decays with  $L_{xy} < 100$  cm occurred within the tracker volume, and the constituent muons are likely to have left tracks in the high-precision tracker. Such events are also subject to high rates of backgrounds from pp collision processes. On the other hand, decays with  $L_{xy} > 300$  cm occurred within the muon system, far away from the interaction point. Tracks reconstructed using the muon system have coarser resolution than tracks reconstructed with the tracker, but are also subject to fewer pp collision backgrounds.

The quantity  $L_{xy}/\sigma_{L_{xy}}$  is referred to as the  $L_{xy}$  significance, as it is a measure of the significance of the fit compared to zero displacement. For a well-reconstructed dimuon corresponding to a long-lived particle decaying a few tens of centimeters away from the PV, as in signal events, this quantity is large. On the other hand, background events from reconstruction mistakes or promptly produced particles, such as can arise in Drell-Yan events, are expected to result in dimuons with small  $L_{xy}$  significance. A large  $L_{xy}$  significance is therefore a strong indicator of consistency with the long-lived particle hypothesis.

The collinearity angle in the transverse plane between the dimuon  $p_T$  vector ( $p_T^{\mu\mu}$ ) and the  $L_{xy}$  vector is referred to as  $|\Delta\Phi|$ . For a real pair of muons originating from a CV formed from the decay of a long-lived particle originating at the PV, the resulting dimuon momentum

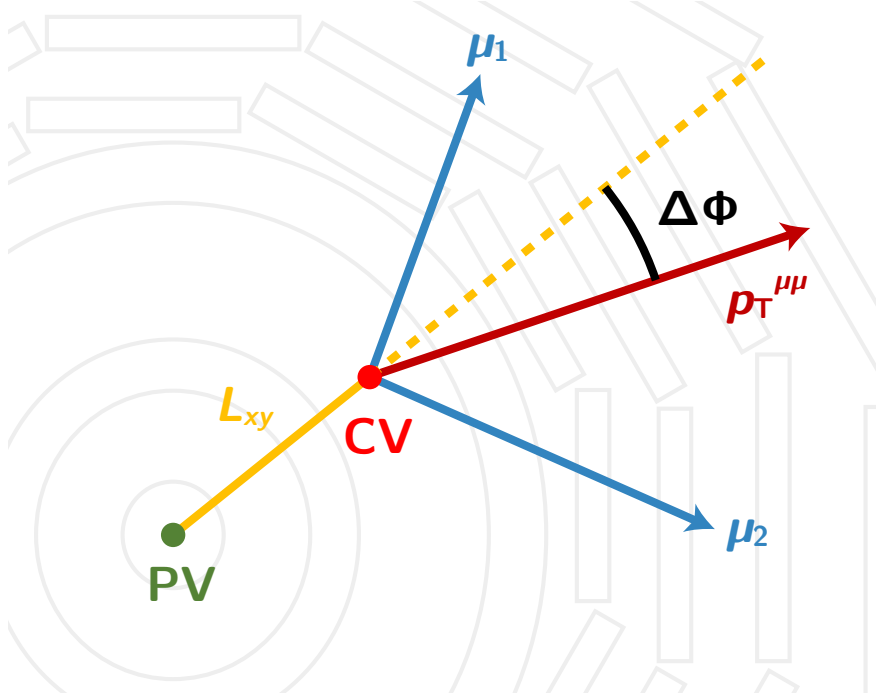


Figure 4.1: Diagram of the key variables used in this analysis. A long-lived particle is produced at the primary vertex (PV) and travels some distance in the detector before decaying into two muons ( $\mu_1$  and  $\mu_2$ ), forming a common vertex (CV). The distance between these two vertices in the transverse plane is referred to as  $L_{xy}$ , the transverse decay length. The angle between the  $L_{xy}$  vector and the  $p_T$  vector of the dimuon system is referred to as  $|\Delta\Phi|$ , the collinearity angle in the transverse plane.

vector points towards the PV, and  $|\Delta\Phi|$  is small. On the other hand, in most types of background events, this quantity is expected to be symmetric around  $\pi/2$ . A small  $|\Delta\Phi|$  is therefore a strong indicator of consistency with the long-lived particle hypothesis.

We follow previous displaced dimuon searches [1, 2, 3] in defining  $L_{xy}$  as the unsigned magnitude of this distance. The sign information regarding which side of the PV the CV is reconstructed is encoded in the companion variable  $|\Delta\Phi|$ , specifically in the sign of  $\cos(|\Delta\Phi|)$ . (We call attention to this fact since other analyses in high-energy physics, for example  $b$  physics analyses, typically attach the sign to  $L_{xy}$ .)

Figure 4.1 illustrates these important quantities and their relationship to each other.

### 4.3 Muon Reconstruction

This analysis is a search for pairs of muons consistent with originating from a common vertex that is displaced with respect to the beam spot. Such dimuon vertices can arise from decays of exotic long-lived particles, which are produced promptly (as a direct result of the hard interaction process and produced close to the beam spot) and travel for some distance before decaying. As mentioned in Section 4.1.1, muons produced from such decays may not point towards the beam spot. Most algorithms performing the offline reconstruction of muon tracks, however, include the beam spot; these reconstructions are described as being “updated at the vertex.” A search for displaced dimuons must therefore use a reconstruction algorithm that does not use this beam spot constraint.

Muons reconstructed using only hits in the muon system are known as standalone muons. Standalone muons are independent of the more precise information from the silicon tracker, and so have poorer spatial and momentum resolution than muons reconstructed using the tracker. However, the reconstruction efficiency for tracker tracks with transverse impact parameter  $d_0$  (the closest distance between the track and the beam spot in the transverse plane) of more than a few tens of cm is essentially zero, while the muon system gives non-vanishing reconstruction efficiency even a few meters away from the beam spot. Standalone muon reconstruction (as well as its variants, discussed below) is available with and without

the beam spot constraint; the reconstructions without the beam spot constraint for more suitable for studying displaced muons. Therefore, variants of standalone muons reconstructed without the beam spot constraint are the starting points for choosing a muon reconstruction suitable for a search for displaced dimuons sensitive to longer lifetimes.

Muon reconstruction in CMS begins with local reconstruction of hits in muon chambers and of segments formed from linear fits to groups of hits within a station [48]. Standalone muon reconstruction follows local reconstruction of hits and segments with a trajectory seeding step and a trajectory building step. A trajectory seed consists of a state vector (track position, momentum, and errors) associated with one or a few hits or segments in the DTs or CSCs. A navigation algorithm uses these trajectory seeds to determine possible layers through which the trajectory may pass, known as compatible layers. A trajectory building step, using these seeds and their compatible layers as inputs, employs a Kalman filter algorithm to iterate over compatible layers and incrementally build a muon trajectory from multiple hits and segments.

Standard trajectory seed generation involves an estimate of the state vector with a parametrization that includes the pp interaction point, and is therefore suitable for reconstruction of muons originating from pp collisions. Another implementation, known as cosmic muon seed generation, modifies the standard algorithm in a few ways:

- Each seed is built from exactly one DT or CSC segment;
- Segments located with a larger  $y$  value in the CMS reference coordinates (Section 3.2.2) are preferred to segments with smaller  $y$  values;
- The trajectory seed state vector direction is determined by the segment direction, excluding the parametrization involving the interaction point;
- The momentum direction for seeds is set to be downward.

These seeds are thus more suitable for reconstruction of muons originating from cosmic rays. As cosmic muons may have large impact parameters, such a reconstruction is also expected to be more suitable for reconstruction of muons originating from highly displaced decays.



The analyses performed using Run 1 data used refitted standalone muons (RSA muons) as their primary muon reconstruction [2, 49]. Compared to ordinary standalone muons without the beam spot constraint (with standard seeding), the RSA muon algorithm computes a final refit (an additional iteration in the standalone muon trajectory builder) excluding the beam spot. This reconstruction was intended to reduce the inherent bias towards the beam spot exhibited by the standard standalone muon reconstruction designed to study muons originating directly from the pp collision. RSA muons thus have improved  $d_0$  and  $p_T$  resolution compared to standalone muons.

Another muon reconstruction algorithm is the displaced standalone muon (DSA muon) reconstruction. Compared to ordinary standalone muons without the beam spot constraint (with standard seeding), the DSA muon reconstruction uses cosmic muon seeding with the modification that the seeds are not forced to point downwards [48, 50]. DSA muon reconstruction was intended to further improve the  $p_T$  resolution of muons produced from highly displaced decays. Since this reconstruction is based on cosmic muon seeds, it is more suitable for reconstruction of displaced muons than the ordinary standalone muon reconstruction.

To compare DSA and RSA muon reconstruction, Figure 4.2 shows graphs of the efficiency as a function of generated  $L_{xy}$  to reconstruct a generated muon as DSA or as RSA, with and without the trigger applied, with respect to  $H \rightarrow 2X \rightarrow 2\mu$  signal events in which both generated muons are within a detector acceptance, defined as

- both generated muon  $p_T > 25 \text{ GeV}$ ;
- both generated muon  $|\eta| < 2$ ;
- generated  $L_{xy} < 500 \text{ cm}$ .

In order to have a sufficient sample size, all 33 of the  $2\mu$  signal samples are combined together for these graphs. The reconstruction efficiency decreases steadily as a function of generated  $L_{xy}$ , dropping off sharply at about 600 cm, corresponding roughly to the start of the third barrel muon station, and the DSA reconstruction is consistently higher than the RSA reconstruction. With the trigger applied, the reconstruction efficiency is high for both

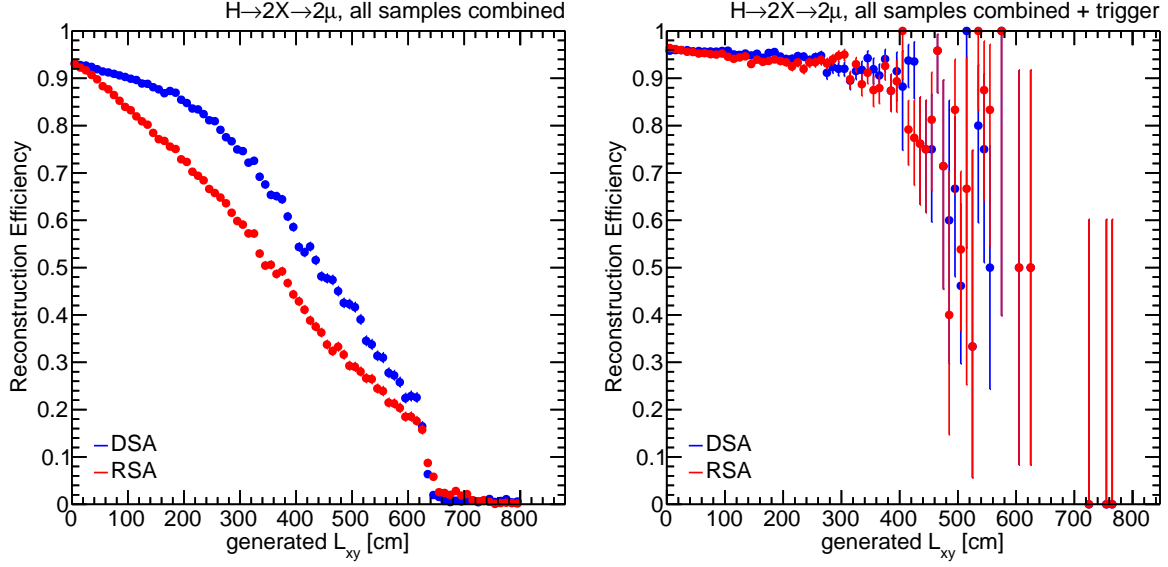


Figure 4.2: DSA and RSA reconstruction efficiency as a function of generated  $L_{xy}$  for  $2\mu$  signal events combining all 33 signal samples, with respect to events within acceptance, (left) without and (right) with the trigger applied. Error bars are for statistical uncertainty only.

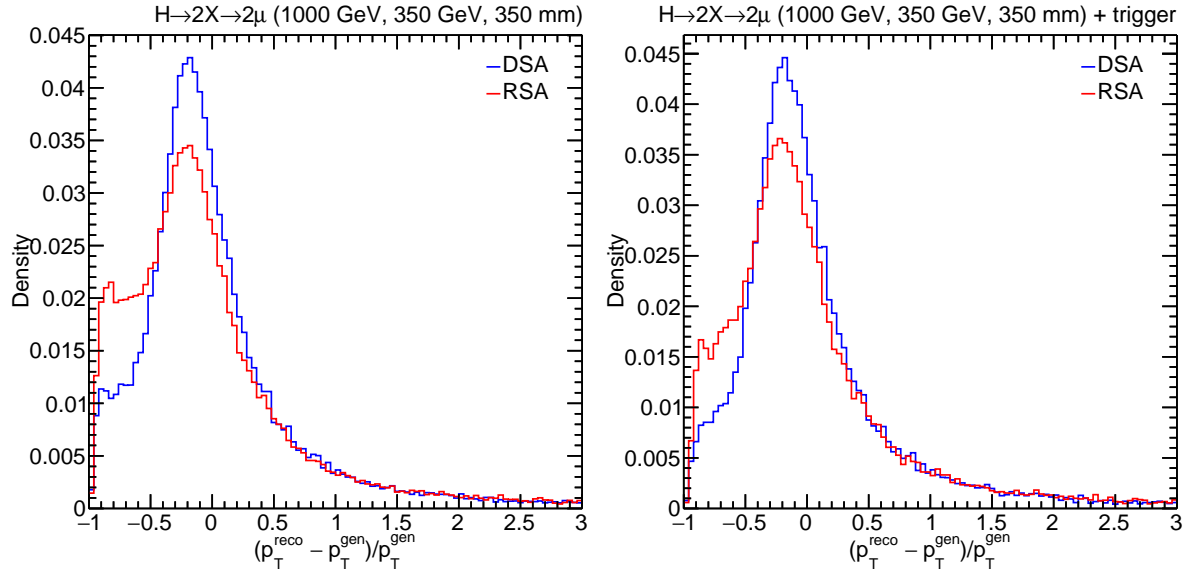


Figure 4.3: DSA and RSA  $p_T$  resolution for generated events within acceptance, (left) without the trigger applied and (right) with the trigger applied, for the  $2\mu$  signal sample with  $m_H = 1000$  GeV,  $m_X = 350$  GeV, and  $c\tau = 350$  mm.

DSA and RSA muons, up until about 300 cm at which point the trigger efficiency seriously limits the number of passing events.

Figure 4.3 shows distributions of the  $p_T$  resolution for the DSA and RSA reconstructions, normalized to unit area, for generated events within acceptance, with and without the trigger applied, for the  $2\mu$  signal sample with  $m_H = 1000$  GeV,  $m_X = 350$  GeV, and  $c\tau = 350$  mm. The RSA reconstruction gives a visibly worse  $p_T$  resolution, with a larger shoulder near  $-1$ , indicating that RSA muons tend to underestimate the muon  $p_T$  compared to DSA muons.

Figures 4.2–4.3 show that DSA muons have both a higher reconstruction efficiency and an improved  $p_T$  resolution over RSA muons. The analysis presented in this thesis therefore primarily uses DSA muons.

## 4.4 Event and Object Selection

### 4.4.1 Overview

This section presents details of the selection criteria used to perform the analysis. Informally, selection criteria are referred to as cuts, as application of selection criteria cuts away events. An overview of the general strategy for selecting reconstructed events and objects consistent with displaced dimuons is as follows; subsequent subsections describe each step in more detail.

1. **DSA muon preselection.** Begin with DSA muons as seeds for searching for displaced dimuons and apply a set of quality criteria to select muons that are reasonably well reconstructed. (Section 4.4.2)
2. **HLT-RECO matching.** Keep only events that contain offline reconstructed muons that match the online trigger muons passing the HLT. (Section 4.4.3)
3. **DSA to PAT muon association.** Attempt to associate DSA muons with global or arbitrated tracker muons (referred to as PAT muons, named after the **Physics Analysis Tools** in the CMS software that process them), and if they can be, replace the DSA muons with the associated PAT muons. This analysis focuses on the DSA muons that are *not* associated with PAT muons, so this step essentially sets aside DSA muons that are consistent with muons reconstructed using the tracker. (Section 4.4.4)
4. **DSA muon object selection.** Apply additional selections to remaining DSA muons that identify DSA muons in signal-like events and ensure that they are of good quality. (Section 4.4.5)
5. **Dimuon formation from common vertex fit.** Form dimuons by fitting two DSA muon tracks to a common vertex. (Section 4.4.6)
6. **Pairing criteria.** Select the 1–2 dimuons most likely to be signal dimuons according to a set of muon pairing criteria. (Section 4.4.7)
7. **Dimuon object selection.** Apply additional selections to dimuons that identify dimuons in signal-like events and ensure they are of good quality. (Section 4.4.8)
8. **Dimuon signal selection.** Apply additional selections to dimuons that require the dimuon be displaced from the beam spot and have a form consistent with the signal hypothesis that the dimuon originates from the decay of a long-lived particle into two oppositely-charged muons. (Section 4.4.9)
9. **Cosmic muon suppression.** Apply additional requirements to suppress background events from cosmic muons. (Section 4.4.10)

#### 4.4.2 DSA Muon Preselection

Event and object preselection begins with DSA muons. The HLT path described in Section 4.1.1 reconstructs muons at Level 2 using a procedure similar to that used to reconstruct muons as refitted standalone (RSA) muons at the offline reconstructed level. However, as discussed in the previous section, DSA muons are chosen as the starting point over RSA muons for their improved  $p_T$  resolution and higher reconstruction efficiency. DSA muons are required to satisfy the following quality criteria:

- reconstructed with hits in at least two muon stations, i.e.

$$N(\text{CSC stations}) + N(\text{DT stations}) > 1$$

- reconstructed with at least 13 hits in the muon stations, i.e.

$$N(\text{CSC hits}) + N(\text{DT hits}) > 12$$

- relative  $p_T$  uncertainty of less than 100%, i.e.

$$\sigma_{p_T}/p_T < 1$$

These quality criteria ensure that the DSA muons have acceptable  $p_T$  resolution and charge assignment. Figure 4.4 shows distributions of the DSA muon  $p_T$  resolution for the  $2\mu$  signal sample with  $m_H = 1000$  GeV,  $m_X = 350$  GeV, and  $c\tau = 350$  mm for events split up by the  $N_{\text{hits}}$  cut and for events split up by the  $\sigma_{p_T}/p_T$  cut, each population normalized to unit area. The events satisfying the cut have reasonable  $p_T$  resolution, while the events failing the cut have rather poor  $p_T$  resolution. Figure 4.5 shows distributions of the difference between reconstructed and generated charge for the same signal sample, for events split up by the  $N_{\text{hits}}$  cut and for events split up by the  $\sigma_{p_T}/p_T$  cut, each population normalized to unit area. The events satisfying the cut are assigned the correct charge in more than 90% of events, while the events failing the cut are assigned the correct charge in only around 60% of events.

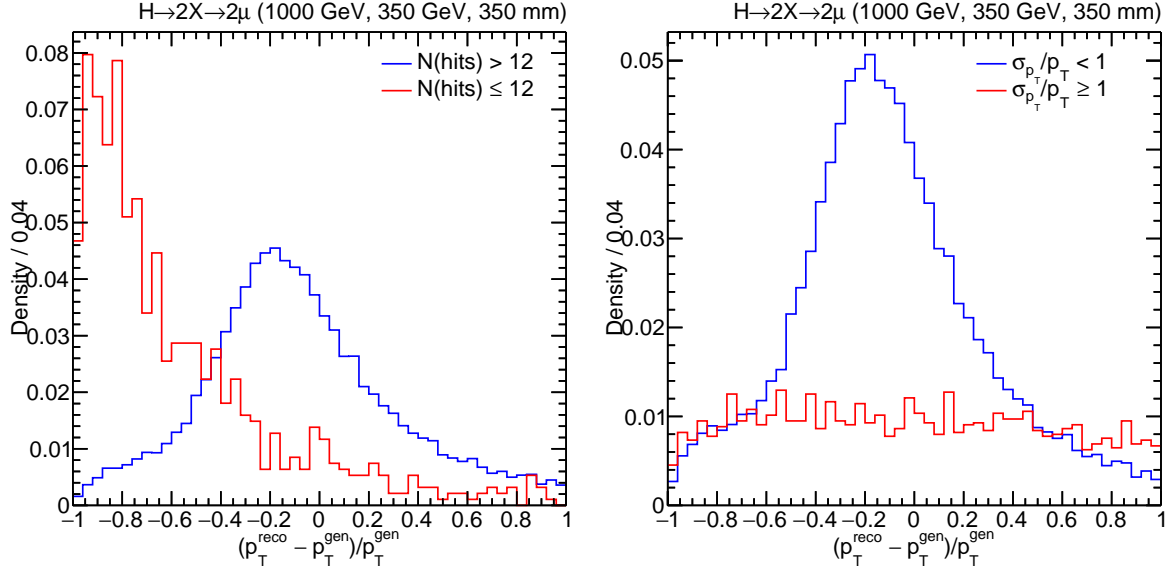


Figure 4.4: DSA muon  $p_T$  resolution, normalized to unit area, for the  $2\mu$  signal sample with  $m_H = 1000$  GeV,  $m_X = 350$  GeV, and  $c\tau = 350$  mm for (left) events with  $N_{\text{hits}} > 12$  and  $N_{\text{hits}} \leq 12$  separately, and (right) events with  $\sigma_{p_T}/p_T \geq 1$  and  $\sigma_{p_T}/p_T < 1$  separately.

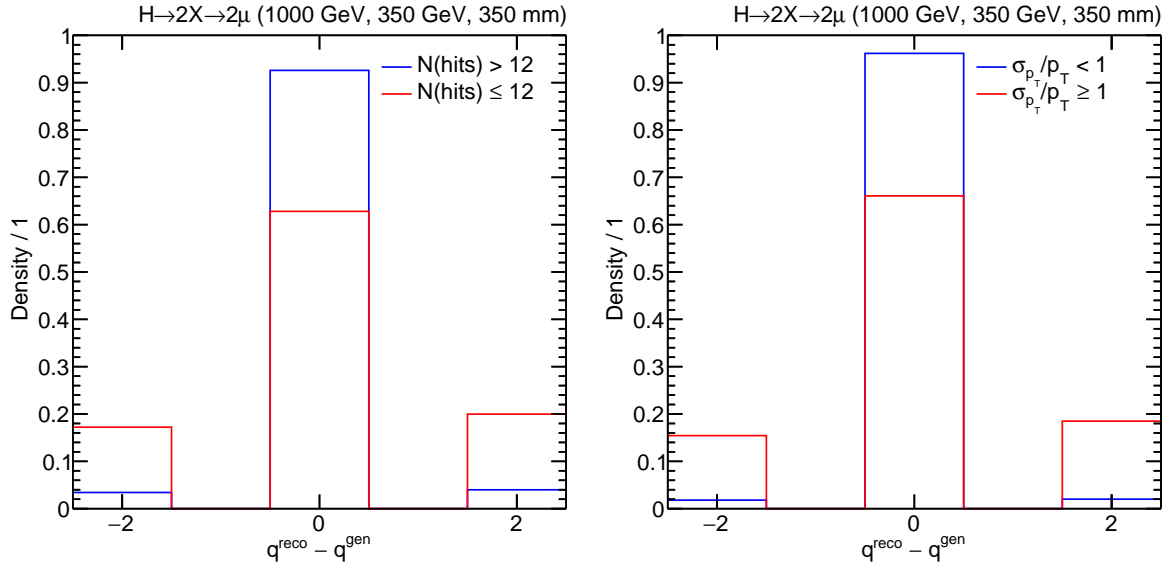


Figure 4.5: DSA muon charge difference, normalized to unit area, for the  $2\mu$  signal sample with  $m_H = 1000$  GeV,  $m_X = 350$  GeV, and  $c\tau = 350$  mm for (left) events with  $N_{\text{hits}} > 12$  and  $N_{\text{hits}} \leq 12$  separately, and (right) events with  $\sigma_{p_T}/p_T \geq 1$  and  $\sigma_{p_T}/p_T < 1$  separately.

### 4.4.3 HLT-RECO Matching

Next, only events in which the HLT decision is matched using the preselected DSA muons are kept, in a procedure referred to as HLT-RECO matching. This ensures that the quickly-reconstructed, coarser-resolution muons reconstructed by the trigger system have counterparts among muons reconstructed offline using the higher-precision data of the full event. For the purposes of HLT-RECO matching, DSA muons are required to pass loose  $p_T$  and  $\eta$  requirements:  $p_T > 10 \text{ GeV}$  and  $|\eta| < 2.0$ . These DSA muons are referred to as DSA muons for trigger matching. Let  $\Delta R$  refer to the angular proximity in  $\eta$ - $\phi$  space between two three-vectors, defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (4.5)$$

Then HLT-RECO matching requires that each of the pair of distinct DSA muons for trigger matching lie within cones of  $\Delta R < 0.4$  around each muon among pairs of muons reconstructed at Level 2 passing the trigger (HLT muons).

The  $p_T$ ,  $\eta$ , and  $\Delta R$  requirements are all chosen to empirically optimize the trade-off between efficiency and purity, matching as many events as possible while keeping the frequency of accidental matches to poor quality muons low. The HLT-RECO matching procedure retains more than 98% of  $2\mu$  signal events that passed the trigger and have at least two DSA muons for trigger matching.

The HLT-RECO matching requirement ensures, at the event level, that the trigger decision is confirmed using objects reconstructed offline. The subsequent object selection, however, does not require that selected dimuons be formed from pairs of triggering HLT muons. This allows the analysis to be sensitive to  $4\mu$  final states, in which two of the four muons passed the trigger, but all four muons are signal muons.

### 4.4.4 DSA to PAT Muon Association

This analysis begins with DSA muons to identify potential displaced muon events, specifically focusing on highly displaced events in which the long-lived particle travels far enough before decaying that a track is not observed in the tracker. Implementing this requirement

necessitates development of a procedure to efficiently associate DSA muons with muons reconstructed using the tracker. This association improves not only potential signal discrimination, but also background rejection. Ensuring that this procedure is both highly efficient (i.e. muons that should be associated, are) and highly pure (i.e. muons that should not be associated, are not) requires some care. This section describes the details of the association procedure.

#### 4.4.4.1 PAT Muons and Matching Criteria

Offline reconstructed muons having either of the following properties are considered:

- **Global muons.** Global muons are muons reconstructed by association a tracker track with a standalone muon by collecting all the hits from both and performing a full track fit, called a global muon fit. Global muons are the primary muon reconstruction used by most CMS analyses.
- **Arbitrated tracker muons.** In order to deal with cases in which global muon reconstruction is inefficient, tracker muons were developed. Tracker muons do not require a standalone muon; instead, a tracker track is extrapolated and associated with hits in the muon system. In the case of multiple tracker tracks associated with the same muon hits, an algorithm called arbitration chooses which association to prefer such that no two tracker muons share muon segments; these tracker muons are referred to as arbitrated tracker muons.

The muons considered in this analysis that are reconstructed offline using both the tracker and the muon system are referred to as PAT muons. (PAT is an abbreviation for **P**hysics **A**nalysis **T**ools, a toolkit that is part of the CMS software and which is used to process event and object reconstruction.) A PAT muon may be both a global muon and an arbitrated tracker muon, and good quality PAT muons are often both.



Candidate PAT muons are associated with DSA muons when they satisfy one of two matching criteria, which are applied in different contexts in a procedure described below. The two matching criteria are:

- **Segment-matched:** A PAT muon is segment-matched to a DSA muon if the PAT muon shares all of its muon system segments with at least 2/3 of the muon system segments used to reconstruct the DSA muon. PAT muon segments and DSA muon segments are compared coordinate by coordinate for equality and are considered the same if their coordinates match.
- **Proximity-matched:** A PAT muon is proximity-matched to a DSA muon if the  $\Delta R$  between
  - the position vector of the innermost hit of the DSA muon and
  - the position vector of the point of closest approach of the PAT muon helically extrapolated (in the magnetic field) to the innermost hit of the DSA muon.

is less than 0.4, with tighter requirements applied in the procedure described below.

The proximity match is defined as it is above because DSA muons are sometimes reconstructed with an incorrect  $\eta$  coordinate, and would not pass a more standard angular proximity match between the momentum directions of the DSA and PAT muons. The helical extrapolation and comparison between the positions improved the efficiency of the proximity match.

#### 4.4.4.2 DSA to PAT Association Procedure

The general procedure is as follows: given a DSA muon, associate segment-matched PAT muons with the DSA muon, using additional criteria to disambiguate among multiple segment-matched PAT muons. These additional disambiguation criteria include requiring that segment-matched PAT muons are also the PAT muon with the smallest proximity-match  $\Delta R$  (as defined above) and/or requiring that segment-matched muons be global and that they be reconstructed from hits in at least 7 tracker layers. If no PAT muons segment-match the DSA

muon, the smallest- $\Delta R$  proximity-matched PAT muon is associated with the DSA muon if its proximity-match  $\Delta R$  is sufficiently small; the threshold is 0.4 if the proximity-matched PAT muon is global and has the same numbers of muon system hits as the DSA muon, and 0.1 otherwise. The full technical details of the procedure are depicted in Figure 4.6.

This procedure prioritizes segment-matched PAT muons over proximity-matched PAT muons. Several combinations of criteria are used to determine the quality selection used to disambiguate segment-matched PAT muons; the combination used here (global and number of tracker layers) is found to be optimal in terms of increasing the efficiency to match to the correct PAT muon. In principle, there exists the case that a proximity-matched PAT muon does not exist or is not among the segment-matched PAT muons. In this case, we were prepared to fall back to associating to the highest- $p_T$  PAT muon among the collection of segment-matched PAT muons. However, this situation never actually occurs in practice.

The  $\Delta R$  threshold in the case that there are no segment-matched PAT muons is chosen empirically to increase the matching efficiency while keeping the rate of false, accidental matches low. At the time of this writing, this analysis does not access the muon system segments used to reconstruct the standalone muons used to form global muons, and so a PAT muon that is global only cannot be segment-matched to a DSA muon. This has little practical effect on the efficiency to correctly associate DSA muons with PAT muons in this analysis. For the small fraction of PAT muons that are global only and should be associated with a DSA muon, the proximity-matched PAT muon is often the correct choice. Therefore, the  $\Delta R$  threshold for proximity-matched global-only PAT muons is set at 0.4, if their numbers of CSC and DT hits and stations are identical to those of the DSA muon. As missed matches can lead to a high rate of background events which are potentially dangerous, this association procedure is also designed to be liberal in order to reject background events: if a DSA muon can be reasonably associated with a PAT muon, it should be, even at the cost of a few accidental matches. Therefore, in the case of no segment-matched PAT muons and a proximity-matched PAT muon that is a tracker muon, the procedure associates with the DSA muon the proximity-matched PAT muon if its proximity  $\Delta R$  is less than 0.1 as a final compromise between efficiency and purity.

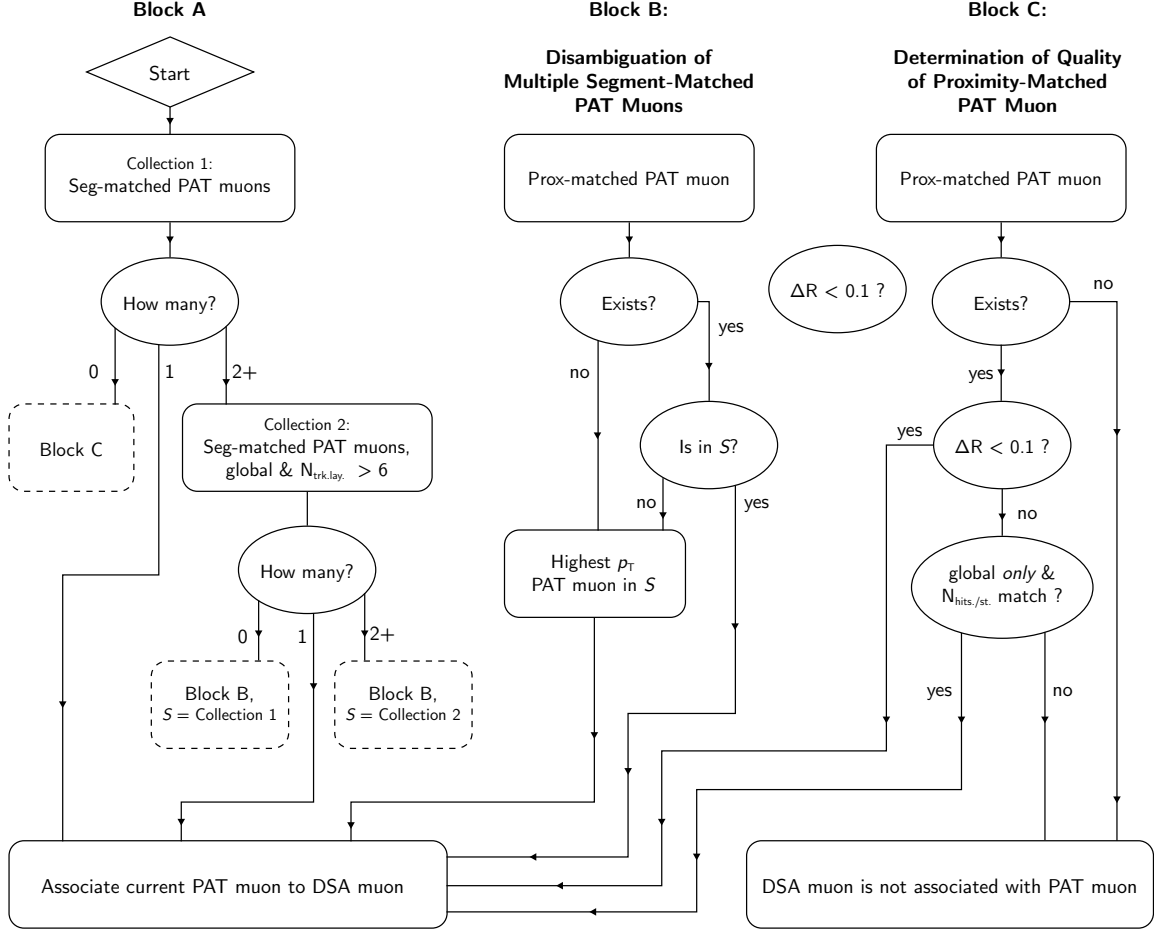


Figure 4.6: Flowchart depicting the technical details of the DSA to PAT association procedure. The procedure begins by considering segment-matched PAT muons. If there are multiple segment-matched PAT muons, then some temporary quality cuts are applied: the requirements that the muon be global and that it be reconstructed from at least 7 tracker layers (abbreviated  $N_{\text{trk. lay.}}$ ). When there are multiple segment matches, the PAT muon that has the smallest proximity-match  $\Delta R$  is used to disambiguate between them, if possible. If there are no segment matches, then the proximity-matched PAT muon is taken as the match if its proximity  $\Delta R$  is sufficiently small. This threshold is 0.4 if the PAT muon is global only (not tracker) and its numbers of CSC and DT hits and stations (abbreviated  $N_{\text{hits./st.}}$ ) matches those of the DSA muon exactly, and 0.1 otherwise.

As mentioned in Section 4.1.3, the analysis presented in this thesis focuses only on the DSA muons that are not associated with PAT muons. To this end, DSA muons are replaced with any associated PAT muons, and so the association step effectively functions as a rejection of DSA muons that can be associated with PAT muons. The effect of this is that this analysis, using DSA muons, does not benefit from the improved resolution of PAT muons, but it benefits tremendously from the superior background rejection. This is the analysis sensitive to the longest lifetimes, with long-lived particles decaying in the outer region of the tracker and outside the tracker through the muon system.

The PAT association procedure is highly effective in suppressing pp collision backgrounds, such as those modeled by the simulated background samples. Figures 4.7–4.8 are histograms of dimuon  $L_{xy}/\sigma_{L_{xy}}$  and  $L_{xy}$  in simulated background samples, with each process scaled to its equivalent luminosity as discussed in Section 4.1.4, before and after the association procedure. The PAT association suppresses events in simulated background by a factor of  $10^5$ , reducing the number of dimuons from 12.5 million to 204. For these histograms, a  $p_T$  cut of 10 GeV is applied to DSA muons on top of the preselection criteria. This is a requirement that is part of the muon object selection (see Section 4.4.5), but is applied here to more clearly demonstrate the effect of the PAT association procedure by suppressing secondary, low- $p_T$  muons not associated with PAT muons, such as those from other pp collisions in the bunch crossing.

In contrast, the association procedure essentially leaves true signal events with dimuon decays outside the tracker relatively untouched. Figure 4.9 is a graph of the number of dimuons (matched to generated signal muons by choosing the closest DSA muons to each generated muon in a cone of  $\Delta R < 0.2$  between their momentum directions) remaining after PAT association (i.e. composed of DSA muons not associated with PAT muons), divided by the number of dimuons before PAT association, as a function of generated  $L_{xy}$ . In order to have a sufficient sample size, all 33 of the  $2\mu$  signal samples are combined together for these graphs. The association efficiently rejects signal events with decays inside the tracker volume,  $L_{xy} < 65$  cm, while rejecting no more than 10% of signal events with decays outside the tracker volume,  $L_{xy} > 65$  cm.

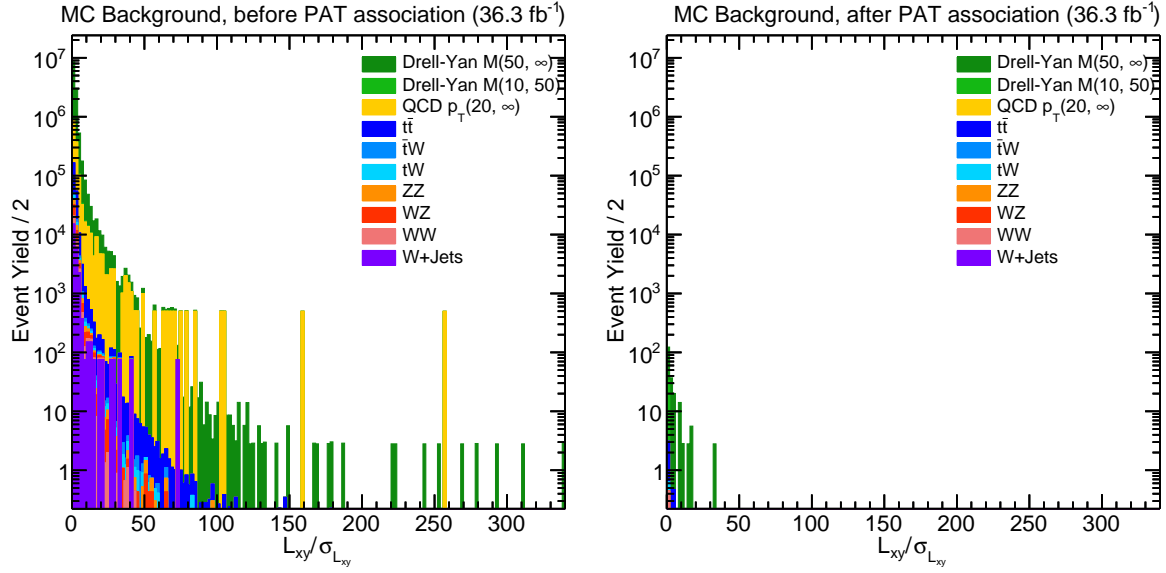


Figure 4.7: Histogram of dimuon  $L_{xy}/\sigma_{L_{xy}}$  in simulated background samples, scaled to 2016 integrated luminosity, (left) before and (right) after the PAT association procedure. The procedure suppresses the number of background events by a factor of  $10^5$ .

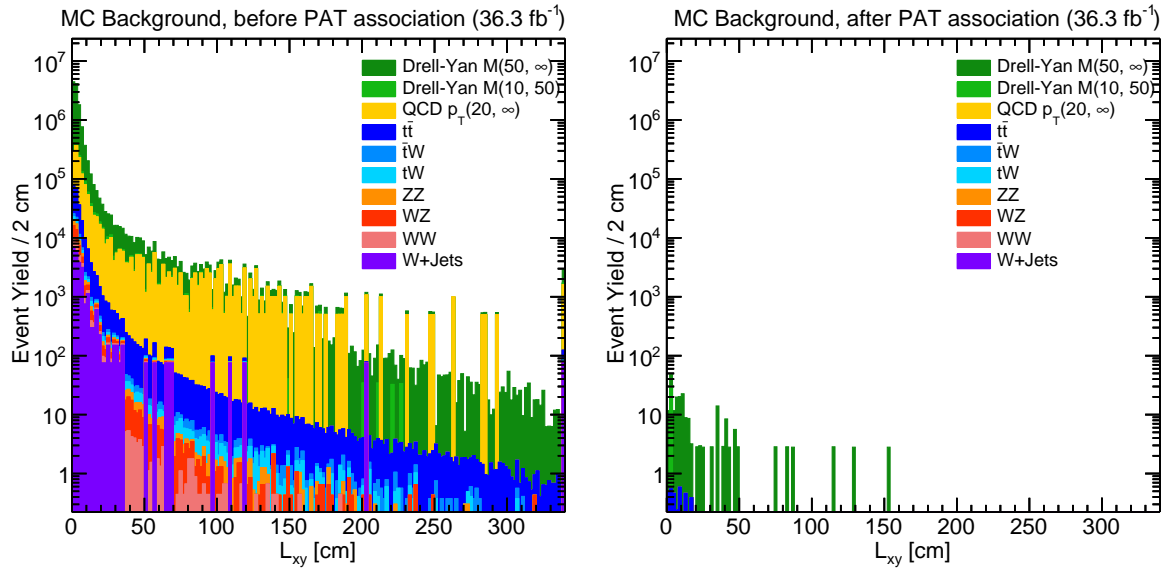


Figure 4.8: Histogram of dimuon  $L_{xy}$  in simulated background samples, scaled to 2016 integrated luminosity, (left) before and (right) after the PAT association procedure. The procedure suppresses the number of background events by a factor of  $10^5$ .

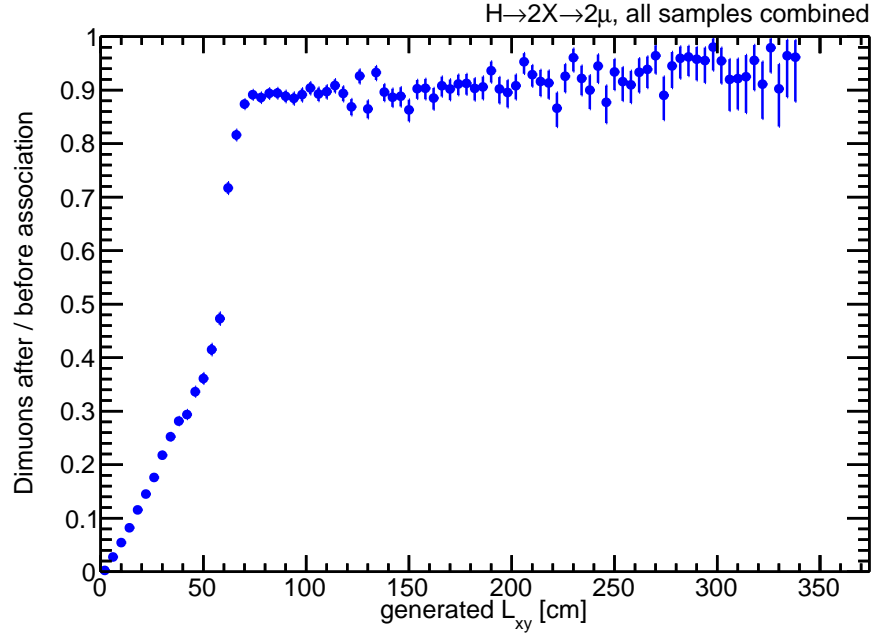


Figure 4.9: Graph of the number of dimuons after the PAT association divided by the number of dimuons before the PAT association, as a function of generated  $L_{xy}$ , in all  $2\mu$  signal samples combined. This graph shows that the association performs as expected, primarily rejecting signal events whose decays occurred within the tracker volume, while accidentally replacing no more than 10% of events outside the tracker volume.

#### 4.4.5 DSA Muon Object Selection

After the DSA to PAT association step explained in the previous section, the analysis considers DSA muons not associated with any PAT muons. As DSA muons are not used by many CMS analyses, a standard set of selections to identify DSA muons does not exist. In order to further select high-quality DSA muons as well as to discriminate signal-like events from background-like events, the following requirements, along with the DSA muon preselection cuts explained in Section 4.4.2, serve as the DSA muon identification selection. DSA muons are required to have

- muon  $p_T$  of at least 10 GeV, i.e.

$$p_T > 10 \text{ GeV}$$

;

- $\chi^2/\text{dof}$  of the muon track fit of at most 2.5, i.e.

$$\chi_{\text{track}}^2/\text{dof} < 2.5$$

- at least 19 hits in the DTs for muons reconstructed only in the barrel, i.e.

$$N(\text{CSC hits}) = 0 \implies N(\text{DT hits}) > 18$$

- time with respect to bunch crossing of at most 12 ns, i.e.

$$|t_{\text{in-out}}| < 12 \text{ ns}$$

The  $p_T$  cut suppresses background events, which often have poor quality muons with low  $p_T$ , including background events arising from QCD processes. The track  $\chi^2/\text{dof}$  cut ensures that the muons are reasonably well reconstructed. The  $N(\text{DT hits})$  cut discriminates signal events from background events. The timing cut is explained in Section 4.4.5.1.

#### 4.4.5.1 In-Time with Triggering Bunch Crossing Requirement

A potentially pernicious class of events that can mimic displaced dimuon decays arises from a combination of technical features in the trigger and readout electronics of the muon chambers and the silicon tracker. In a normal event with two muons, the timing of the trigger pulse from the two muons in the muon chambers is correctly aligned in time with the hit readout of the muon chambers and the tracker. It can happen, however, that jitter in the muon trigger electronics creates a Level-1 trigger signal that is 25 ns earlier than normal. In the muon system readout, the muon hits are still recorded, but they have hit times that are 25 ns later than normal (since they are measured with respect to the Level-1 trigger signal that is 25 ns too early). When there is an early Level-1 trigger signal, in contrast, the tracker hits associated with the triggering muons are *not* recorded; tracker hits from unrelated pp collisions 25 ns earlier are recorded instead. The offline reconstruction code then finds the two triggering muons, but not the associated tracker tracks. This mimics the main feature of displaced muons: tracks in the muon system with no associated tracker tracks.

Fortunately, the recorded times of the hits in the muon system provide a means to recognize such (very rare) events. In practice, the time of the DSA muons is obtained from a precision timing algorithm that is implemented for standard standalone (SA) muons. This algorithm provides a time that is centered on zero for normal muons, with jitter of a few nanoseconds, computed by extrapolating to the interaction point the times each constituent hit was formed with respect to the bunch crossing, under the hypothesis that the muon was produced in the detector and traveled outwards. This time is referred to as  $t_{\text{in-out}}$ .

Matching DSA muons to SA muons is straightforward: DSA muons with an SA muon within a  $\Delta R$  cone of 0.2 are considered matched to the SA muon. The time associated with the SA muon can be used to distinguish between normal muons and the pathological cases where the times are centered on non-zero multiples of 25 ns (the time between bunch crossings). This requirement is implemented as the requirement  $|t_{\text{in-out}}| < 12 \text{ ns}$ , and is more than 99% efficient for simulated samples of the  $2\mu$  signal.



This class of background was understood only after unblinding the signal region, and so this selection criterion was added with the knowledge that it would eliminate some events. However, we believe that this is a clear case that any bias due to adding this criterion is negligible, and that it would be foolish not to add this criterion, given our current understanding of this background.

#### 4.4.6 Dimuon Formation from Common Vertex Fit

A decay of a long-lived particle to two muons is detected as a pair of muons originating from a common vertex, so at this stage, pairs of reconstructed muons are investigated together for consistency with originating from a common vertex and from the decay of a massive, long-lived particle. A pair of DSA muon tracks fit to a common vertex, along with the four-momentum sum of the two muons, is referred to as a dimuon. All  $n(n - 1)/2$  possible pairs of distinct DSA muons among  $n$  selected DSA muons are initially considered when forming dimuons. This set of pairs is immediately filtered by requiring that the distance of closest approach (DCA) of the two DSA tracks helically extrapolated in the magnetic field is less than 50 cm:

$$\text{DCA} < 50 \text{ cm}$$

This is a loose requirement ensuring that the common vertex fit is not performed on a pair of tracks that never come close to approaching each other.

##### 4.4.6.1 Vertex Fitting

The common vertex fit is performed by an implementation of the Kalman filter algorithm in the CMS software, called the `KalmanVertexFitter` [51, 52]. The default implementation in the CMS software version used in this analysis restricts the range of the location of the fitted vertex to approximately within the boundary of the silicon tracker. In this analysis, this restricted range is extended to the beginning of the muon system in order to efficiently reconstruct a common vertex for particles decaying outside the tracker.

The technical changes to the CMS software are documented here. The following changes were made to `RecoVertex/VertexTools/src/SequentialVertexFitter.cc`:

- `TrackerBoundsRadius` was changed from 112 to 500;
- `TrackerBoundsHalfLength` was changed from 273.5 to 1000.

The results of the common vertex fit are used to define dimuon quantities such as transverse decay length. As the existence of a common vertex fit is necessary to further study dimuon properties and consistency, the common vertex fit is required to converge. The common vertex fit modifies the input tracks in order to be consistent with originating from a common vertex; these tracks are referred to as “vertex-constrained”, and any of the muon properties ( $p_T$ , direction, etc.) may be different in the vertex-constrained version of the track than in the original version. For many purposes, the vertex-constrained tracks are preferable because they represent a reconstruction performed with more information. However, in a few situations, the vertex fitter fails or produces nonsensical results. A few of the features of the fitter are documented here.

The vertex fit does not always converge for an arbitrary pair of tracks. The left plot of Figure 4.10 is a graph of the efficiency for the common vertex fit to converge, with respect to signal events in which both generated muons are reconstructed as DSA muons, as a function of the generated  $L_{xy}$ , for the  $2\mu$  signal sample with  $m_H = 400$  GeV,  $m_X = 150$  GeV, and  $c\tau = 4000$  mm.

The denominator of this efficiency is the number of generated events within acceptance (both generated muon  $p_T > 25$  GeV, both generated muon  $|\eta| < 2$ , and generated  $L_{xy} < 500$  cm) in which both generated muons had matching DSA muons. That is, each generated muon has a DSA muon passing the muon preselection and the muon object selection within a cone of  $\Delta R < 0.2$  between their momentum directions. The numerator of this efficiency is the number of such events in which the dimuon common vertex fit converged for the two matched muons. In order to observe the effect of the common vertex fit independently of the low trigger efficiency at large  $L_{xy}$ , events are not required to pass the trigger

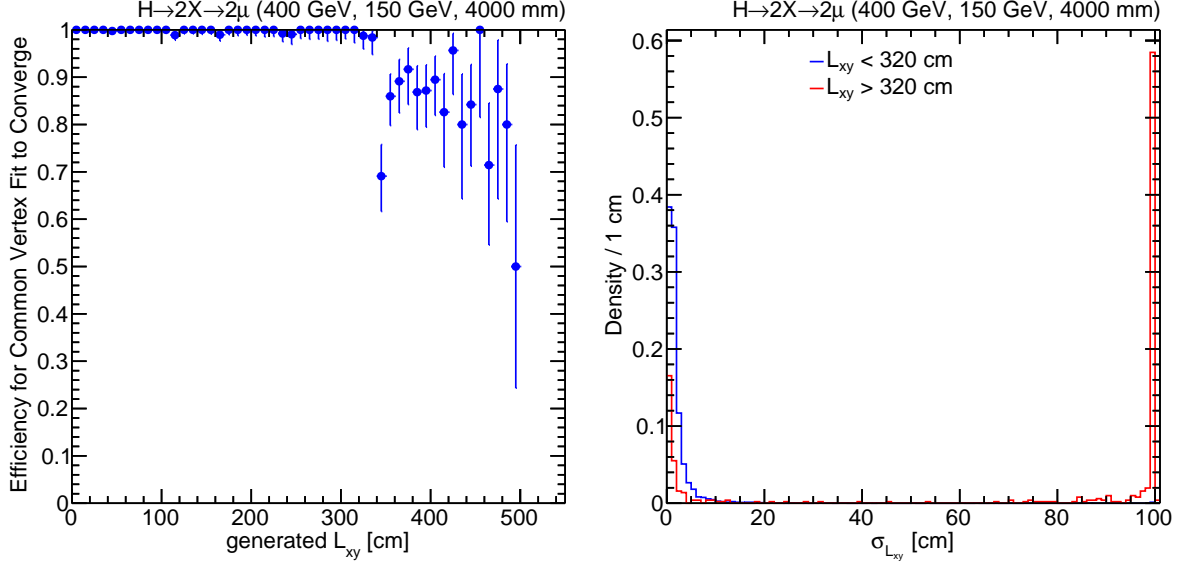


Figure 4.10: (Left) Efficiency for the common vertex fit to converge for the  $2\mu$  signal sample with  $m_H = 400$  GeV,  $m_X = 150$  GeV, and  $c\tau = 4000$  mm as a function of generated  $L_{xy}$ , for generated events within acceptance. Both the preselection and object selection cuts are applied to DSA muons. In this graph, events are not required to pass the trigger, the HLT-RECO matching requirement was dropped, and the DSA to PAT association procedure was not performed. Error bars are for the statistical uncertainty only. (Right) Histograms of  $\sigma_{L_{xy}}$  normalized to unit area for the events in the numerator of the left plot, separately for  $L_{xy} < 320$  cm and  $L_{xy} > 320$  cm, for the  $2\mu$  signal sample with  $m_H = 400$  GeV,  $m_X = 150$  GeV, and  $c\tau = 4000$  mm, for generated events within acceptance. The distribution for  $L_{xy} > 320$  cm has a large peak near 100 cm.

and the HLT-RECO matching requirement was dropped. In order to have a sufficient sample size, the DSA to PAT association procedure was not performed.

Convergence of the vertex fit is highly efficient when fitting two DSA muon tracks that are matched to generated signal muons produced from long-lived particle decaying in the detector up to transverse displacements of 330 cm. Beyond 330 cm, the efficiency for the common vertex fit to converge drops dramatically.

For those events beyond 330 cm that do converge, the fit quality is often quite poor, e.g. the  $p_T$  resolutions of the vertex-constrained tracks are worse than before the fit. Finally, many of these events retain an internal  $L_{xy}$  uncertainty ( $\sigma_{L_{xy}}$ ) of a default value of 100 cm. The right plot of Figure 4.10 shows distributions, normalized to unit area, of the events in the numerator of the left plot of Figure 4.10, separately for  $L_{xy} < 320$  cm and  $L_{xy} > 320$  cm. A large fraction of the events with  $L_{xy} > 320$  cm have  $\sigma_{L_{xy}}$  near 100 cm. Such events would have a small  $L_{xy}$  significance (3 or less) and would not pass an analysis selection. This type of common vertex fit failure would result in the loss of these events, even though the fit converged.

As in the Run 1 analyses, the trigger efficiency at such large  $L_{xy}$  values is quite small, and the  $L_{xy}/\sigma_{L_{xy}}$  selection discussed in Section 4.4.9 usually suppresses such events as well. Therefore, further study of these events and attempts to rescue them from the behavior of the vertex fitter are of low priority for this analysis, and the underlying causes for this behavior remain a curiosity.

#### 4.4.7 Pairing Criteria

Considering all possible pairs of DSA muons results in many dimuons formed from DSA muons that are not related in any way. It is therefore important to develop criteria that choose the correct reconstructed dimuons with high efficiency, consistent with signal. Selections on individual dimuons (such as requiring the vertex  $\chi^2$  to be small) impose some quality constraints that are useful for selecting such dimuons, but any set of criteria that determines which pairs of muons form the correct dimuons must consider the event as a

whole. Choosing such a set of pairing criteria is particularly subtle in, for example, decays of two long-lived particles into a final state with 4 muons. The 4 muons can be nominally paired into 6 overlapping dimuons, which can be partitioned into three distinct sets of 2 dimuons with no shared muons; the criteria must choose between them. This situation is made more complex when, among other things,

- one or more muons are not reconstructed;
- duplicate muons are reconstructed from partial sets of hits;
- muons are not all reconstructed with the correct charge;
- pairs of muons are highly collimated and distinguishing them is difficult;
- vertex fits of muons from different long-lived particle decays yield a good vertex  $\chi^2$ .

In developing this set of pairing criteria, several combinations of metrics were considered.

- Dimuon(s) formed from the highest  $p_T$  muons in the event
- Dimuon(s) with the least vertex  $\chi^2$  in the event
- Dimuon(s) formed from muons with charges of opposite sign
- Pairs of dimuons with the least sum of vertex  $\chi^2$
- Pairs of dimuons with the least difference in reconstructed dimuon mass

Figure 4.11 is a diagram illustrating the application of the pairing criteria to the set of all possible dimuons in the case of four selected muons.

Potential pairing criteria were studied and optimized on both the  $2\mu$  and  $4\mu$  simulated signal samples. Choosing the highest  $p_T$  muons in the event is highly correlated with choosing the signal muons, a fact that is robust across all signal sample parameters covering a wide range of long-lived particle lifetimes and muon  $p_T$  spectra. Requiring that dimuons be formed from muons of opposite charge provides only modest efficiency gains with respect to choosing the signal dimuons, and is undesirable at this stage as it introduces dependence on a specific

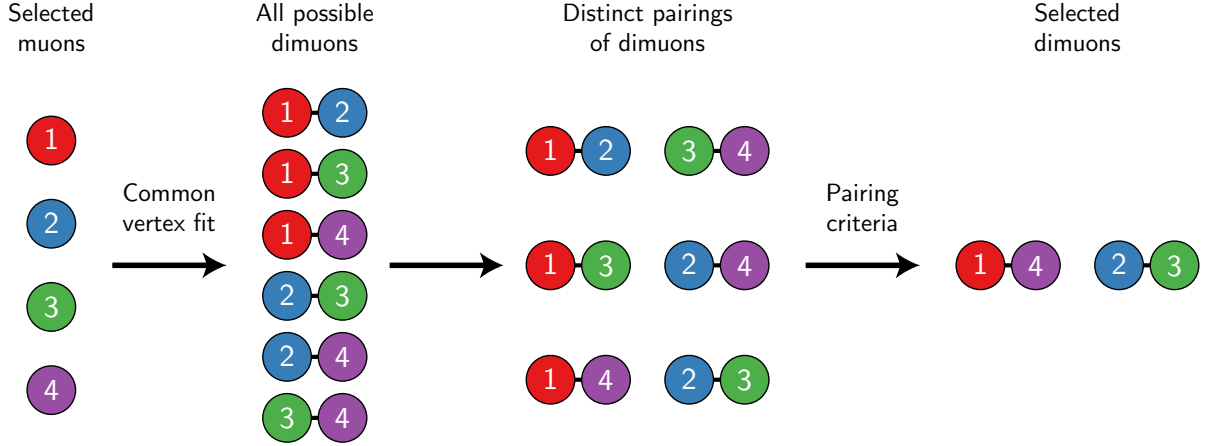


Figure 4.11: Diagram illustrating the application of pairing criteria to dimuons in the case of four selected muons. Four muons may be paired into six dimuons whose constituent muons overlap. These six dimuons may be partitioned into three distinct pairings of dimuons, in which no muons are shared between dimuons. Pairing criteria select the correct pair of dimuons consistent with a four muon signal event.

type of signal model. In events with fewer than four muons, a simple ranking of all possible dimuons by vertex  $\chi^2$  yielded the highest efficiency.

The combinatorial space is far richer in events with four or more muons. For long-lived particles decaying outside the tracker leading to events with at least 4 muons, the criterion yielding the highest efficiency to select signal dimuons is to choose the pair of dimuons whose vertex  $\chi^2$  sum is the smallest, among all distinct pairs of dimuons that can be formed from the 4 highest  $p_T$  muons in the event. An alternative criterion to the least sum of vertex  $\chi^2$  is the least difference in reconstructed dimuon mass. Due to the limited mass resolution of DSA muons, this criterion was found to be less efficient overall than the least sum of vertex  $\chi^2$ , except in events with a generated transverse decay length of less than 30 cm, a region where this DSA muon-based analysis has low sensitivity. Applying the pairing criteria to an event with any number of dimuons formed from any number of muons results in 2 or fewer selected dimuons for the event. The full technical details of this procedure are depicted in Figure 4.12.

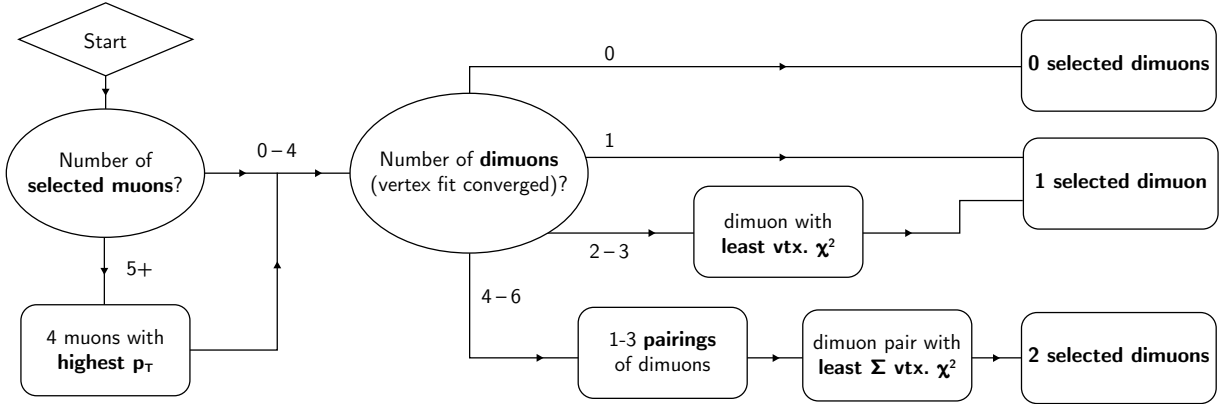


Figure 4.12: Flowchart depicting the technical details of the pairing criteria procedure. Up to 4 DSA muons are selected, ranked by  $p_T$ . With the DCA and vertex fit convergence requirements, these muons can be formed into up to 6 dimuons (0 for 0 or 1 muon, up to 1 for 2 muons, up to 3 for 3 muons, and up to 6 for 4 muons). The pairing criteria choose up to 2 dimuons among the possible dimuons.

Figure 4.13 shows graphs of the efficiency for the pairing criteria to correctly select reconstructed dimuons as a function of generated  $L_{xy}$ , for a selection of  $2\mu$  and  $4\mu$  signal samples. Figure 4.14 shows the same graphs, but for all 33 signal samples combined together. The denominator of this efficiency is the number of generated signal dimuons matching a reconstructed dimuon (using the closest DSA muons to each generated muon in a cone of  $\Delta R < 0.2$  between their momentum directions). The numerator of this efficiency is the number of reconstructed dimuons chosen by the pairing criteria that are the same as the reconstructed dimuon matched to generated signal. This efficiency is 97–100% for  $2\mu$  and 80–98% for  $4\mu$  signal samples with medium and long lifetimes, for generated transverse decay lengths of  $L_{xy} > 100$  cm. The pairing criteria are less efficient for  $4\mu$  samples with larger values of  $m_H$  and smaller values of  $m_X$ , because events in such samples consist of two pairs of highly collimated muons, which yield a greater uncertainty in the fitted vertex position in the dimuon momentum direction, and therefore less discriminating vertex  $\chi^2$  values.

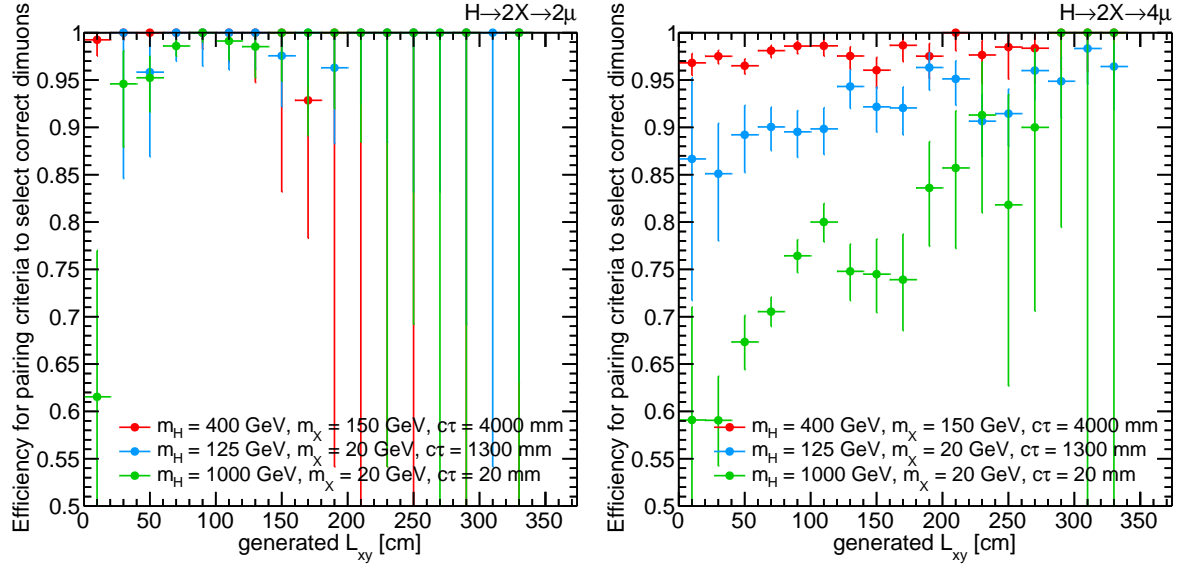


Figure 4.13: Efficiency for the pairing criteria to correctly choose reconstructed dimuons as a function of generated  $L_{xy}$  for (left)  $2\mu$  signal samples and (right)  $4\mu$  signal samples, for selected signal parameters. The behavior varies with signal parameters, and is lower for  $4\mu$  signal samples, but overall the efficiency is high outside the tracker volume ( $L_{xy} > 100$  cm).

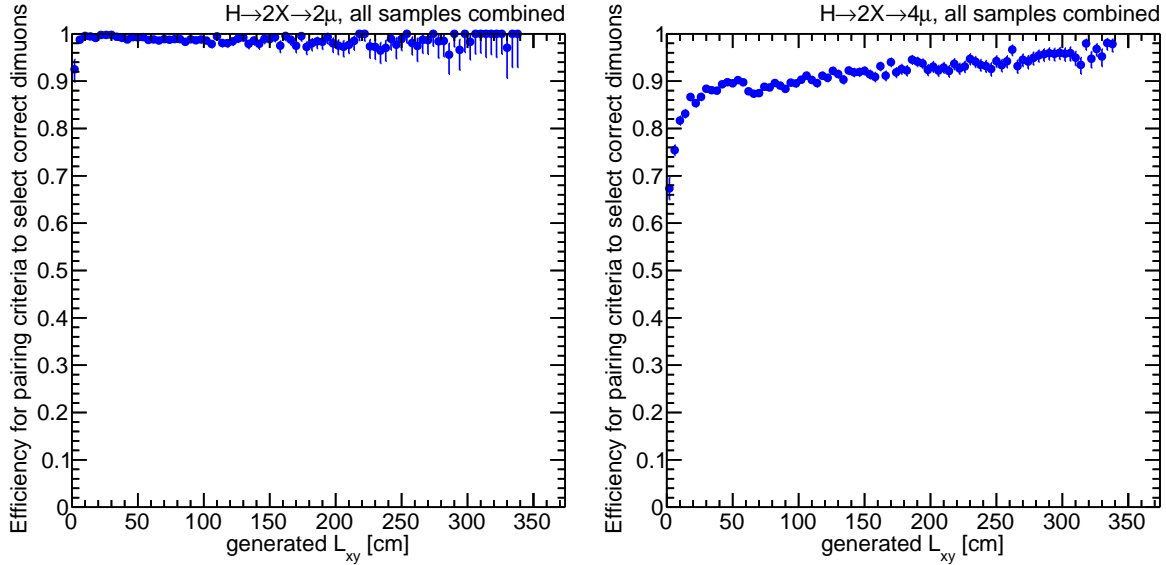


Figure 4.14: Efficiency for the pairing criteria to correctly choose reconstructed dimuons, as in Figure 4.13, for all samples combined.



#### 4.4.8 Dimuon Object Selection

Along with the DCA requirement, the requirement of convergence of the common vertex fit, and the application of the pairing criteria explained in Section 4.4.7, the following requirements serve as a dimuon identification selection. These requirements further select high-quality dimuons as well as suppress background events. Dimuons are required:

- to have a reconstructed dimuon mass of at least 10 GeV, i.e.  $m_{\mu\mu} > 10 \text{ GeV}$ ;
- to have a  $\chi^2$  of the common vertex fit of at most 20, i.e.  $\chi^2_{\text{vertex}} < 20$ .

The vertex  $\chi^2$  cut ensures that the dimuon is formed from tracks that can be reasonably well associated with a common vertex. The mass cut suppresses complex backgrounds arising from QCD processes and events with low  $p_T$  and collimated muons.

#### 4.4.9 Dimuon Signal Selection

The following criteria select displaced dimuons consistent with the signal hypothesis, i.e. that the dimuon was constructed from two muons with opposite-sign charge originating from the decay of a long-lived particle produced promptly, resulting in a dimuon vertex displaced from the beam spot. The relevant quantities were defined in Section 4.2. Dimuons are required:

- to have an  $L_{xy}$  significance of at least 6, i.e.  $L_{xy}/\sigma_{L_{xy}} > 6$ ;
- to have a transverse collinearity angle of less than  $\pi/4$ , i.e.  $|\Delta\Phi| < \pi/4$ ;
- to be formed from two DSA muons with electric charges of opposite sign.

#### 4.4.10 Cosmic Muon Suppression

As mentioned in Section 4.4.6.1, one result of the common vertex fit of two muons is a pair of vertex-constrained tracks. After the common vertex fit, a population of selected dimuons with relatively small vertex  $\chi^2$  values was observed in data, and not in simulation, with the following unusual properties, compared to before the common vertex fit:

- Muon vertex-constrained momentum  $\phi$  directions change by approximately  $\pi$ ;
- Muon vertex-constrained  $p_T$  uncertainties are unusually small, i.e.  $\sigma_{p_T}/p_T \approx 10^{-7}$ – $10^{-5}$ .

The presence of these dimuons are found to correlate with

- Muons with relatively large transverse impact parameters, i.e.  $d_0 \approx 200$ – $1000$  cm, and
- Events with no pp collision vertices, and/or
- Events with large numbers of DSA muons, i.e.  $N(\text{DSA}) \approx 15$ – $20$ ;
- And these large numbers of DSA muons are largely parallel, i.e.  $|\cos \alpha| \approx 1$ .

Further investigation suggested that these events are consistent with multiple cosmic muons arriving simultaneously from a single atmospheric shower. Figure 4.15 is an example display of an event in data consistent with multiple cosmic muons from a single shower in a pp collision event, containing a large number of approximately parallel pairs of DSA muons in addition to two global muons (not selected) originating from the pp collision.

A cosmic muon can be reconstructed as two back-to-back muons, one in the upper half of the detector and one in the lower half of the detector. A reconstructed dimuon could also be formed from half a cosmic muon and a track from the pp collision. These dimuons are reconstructed as highly displaced, which is why the muons have such large  $d_0$  values. The common vertex fit discussed above has anomalous behavior at large displacements, so the presence of the flip-by- $\pi$  and small  $p_T$  uncertainty anomalies when fitting to muons with large  $d_0$  is perhaps not unexpected. Such events are background events and so the analysis imposes a set of selections designed to suppress events with multiple cosmic muons.

Some of these events have no pp collision vertices and contain only cosmic muons. Events are therefore required to contain at least one well-identified vertex with position  $(x, y, z)$  satisfying the following requirements:

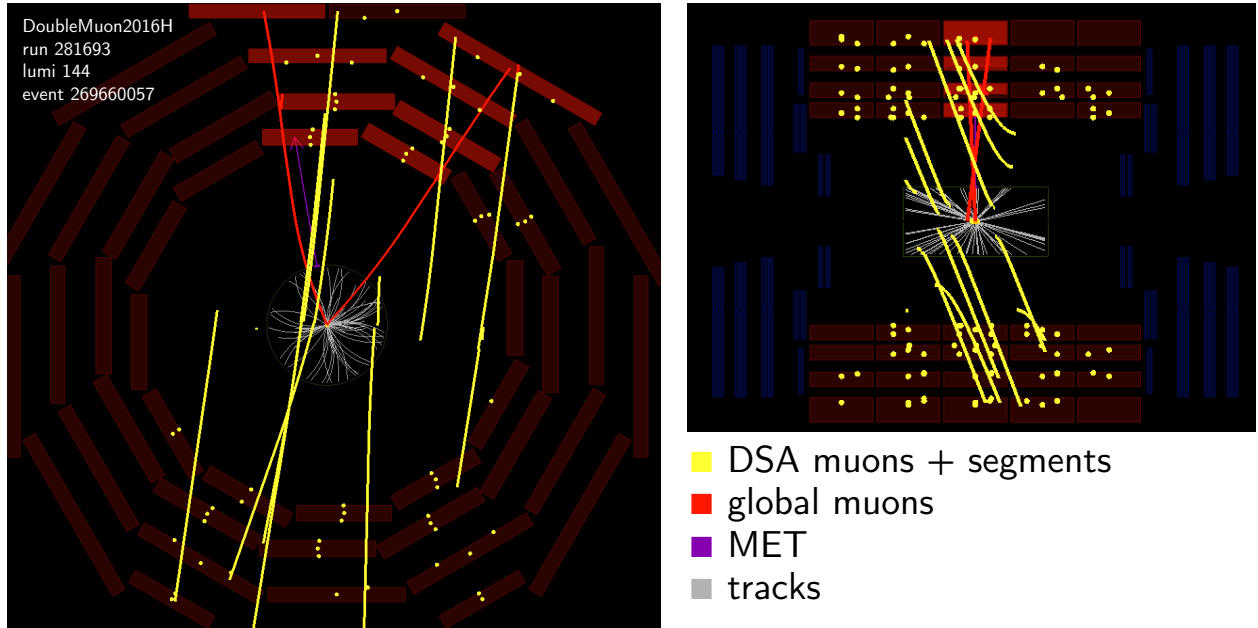


Figure 4.15: Display of an event in data consistent with multiple cosmic muons from a single atmospheric shower in a pp collision event. Two of the many approximately parallel DSA muons (yellow) reconstructed in this event formed a dimuon vertex with a good vertex  $\chi^2$  and the event subsequently passed all our other selections. Requirements that events not contain too many parallel pairs of DSA muons and that the opening angle between the two muons not be too close to  $\pi$  were found to be effective in suppressing such events.

- At least 4 degrees of freedom when constructing the vertex;
- $|z| < 24 \text{ mm}$ ;
- $\sqrt{x^2 + y^2} < 2 \text{ mm}$ .

This set of requirements is referred to in CMS as the **PrimaryVertexFilter**.

A typical strategy to suppress back-to-back cosmic muons is with a selection on  $\cos \alpha$ , and indeed the trigger has just such a cut, roughly equivalent to  $\cos \alpha > -0.8$ . Following the trigger, our offline analysis selection also requires  $\cos \alpha > -0.8$ , on the opening angles both between the vertex-constrained muons and between the original (before the vertex fit) muons.

However, these two requirements alone are not sufficient to suppress all events consistent with multiple cosmic muons from an atmospheric shower. As an additional measure to suppress cosmic muons, the analysis rejects events with a large number of parallel pairs of DSA muons. All possible pairs of DSA muons, with no common vertex fit, are considered, and the number  $N(\text{parallel pairs})$  of such pairs with  $|\cos \alpha| > 0.99$  are counted. The analysis selection then requires that events contain no more than 5 such pairs, verified to be of negligible efficiency in signal.

In summary, the cosmic muon suppression selections are

- Events pass the **PrimaryVertexFilter**;
- Events have  $N(\text{parallel pairs}) < 6$ ;
- Dimuons have original and vertex-constrained  $\cos \alpha > -0.8$ .

#### 4.4.11 Summary of Event and Object Selection

Table 4.4 summarizes the full event and object selections discussed in this section. When setting upper limits, the mass cut is further varied as a function of signal model, as explained in Section 4.6.2.

Event Selection		
primary vertex	<code>PrimaryVertexFilter</code> passed	
HLT-RECO matching	HLT-RECO matching algorithm found match	
number of parallel DSA muon pairs	$N(\text{parallel pairs})$	$< 6$
DSA Muon Selection		
association with PAT muons	DSA muons <i>not</i> associated with PAT muons	
number of CSC and DT stations	$N(\text{CSC+DT stations})$	$> 1$
number of CSC and DT hits	$N(\text{CSC+DT hits})$	$> 12$
number of DT hits for barrel muons	$N(\text{DT hits}   \text{barrel})$	$> 18$
relative $p_T$ uncertainty	$\sigma_{p_T}/p_T$	$< 1$
transverse muon momentum	$p_T$	$> 10 \text{ GeV}$
normalized track $\chi^2/\text{dof}$	$\chi^2_{\text{track}}/\text{dof}$	$< 2.5$
time at interaction point	$ t_{\text{in-out}} $	$< 12 \text{ ns}$
Dimuon Selection		
distance of closest approach of muon	DCA	$< 50 \text{ cm}$
common vertex fit	common vertex fit converged	
pairing criteria	best 1–2 ranked dimuons selected	
dimuon mass	$m_{\mu\mu}$	$> 10 \text{ GeV}^*$
vertex $\chi^2$	$\chi^2_{\text{vertex}}$	$< 20$
cosine of dimuon 3D opening angle	$\cos \alpha$	$> -0.8$
$L_{xy}$ significance	$L_{xy}/\sigma_{L_{xy}}$	$> 6$
transverse collinearity angle	$ \Delta\Phi $	$< \pi/4$
opposite sign muons	constituent muons are oppositely charged	

Table 4.4: Summary of full selection, organized into event, muon, and dimuon requirements. The asterisk refers to a selection that varies depending on the signal model used; see Section 4.6.2.

## 4.5 Background Estimation

As mentioned above, due to MC simulation limitations, evaluating the expected background in the selection is performed using events in data. To avoid potential bias while optimizing the analysis, the events passing the full selection in data (the signal region) were blinded until the last steps of the analysis, and the background was estimated using the numbers of events in data in several control regions, which are selections (in the parameter space of the analysis quantities) of subsets of events that are expected to be free of signal events.

As discussed in Section 4.4.5.1, after unblinding there was a change made to the selection, notably rejection of a clear class of background events with bad timing. We also realized that same-sign events could be put to better use in quantitative estimates of the opposite-sign background from QCD events, beyond the rough qualitative use that we had foreseen.

One type of control region is defined as an interval of the transverse collinearity angle,  $|\Delta\Phi|$ . As discussed in Section 4.2, the signal and background have different distributions in  $|\Delta\Phi|$ . For dimuons consistent with the signal hypothesis, the dimuon momentum vector should be consistent with that of a particle originating from the primary vertex, and so  $|\Delta\Phi|$  is expected to be small, peaking at zero. Figure 4.16 shows the distribution of  $|\Delta\Phi|$  for  $2\mu$  signal events (all 33 sets of signal parameters combined) passing the full selection (as described in the previous section), except for the  $|\Delta\Phi|$  cut.

On the other hand, dimuons in some background events, such as misreconstructed Drell-Yan (DY) events, are expected to have a  $|\Delta\Phi|$  distribution symmetric around  $\pi/2$ , because the dimuon momentum vector is uncorrelated with the direction of flight from the primary vertex. This expected symmetry of  $|\Delta\Phi|$  motivates the definition of the signal region as events with  $|\Delta\Phi|$  near 0, i.e.  $|\Delta\Phi| < \pi/4$  and a symmetric control region as events with  $|\Delta\Phi|$  near  $\pi$ , i.e.  $|\Delta\Phi| > 3\pi/4$ . (The choices of these values are explained in Section 4.5.2.) A transfer factor ( $\tau$ ) is defined as the ratio between the number of background events in the signal region to that in the control region. (Note that this is the reciprocal of the quantity  $\tau$  defined in Reference [53].) With  $\tau = 1$ , the number of events in the control region is a nominal estimate of the expected (Drell-Yan) background in the signal region.

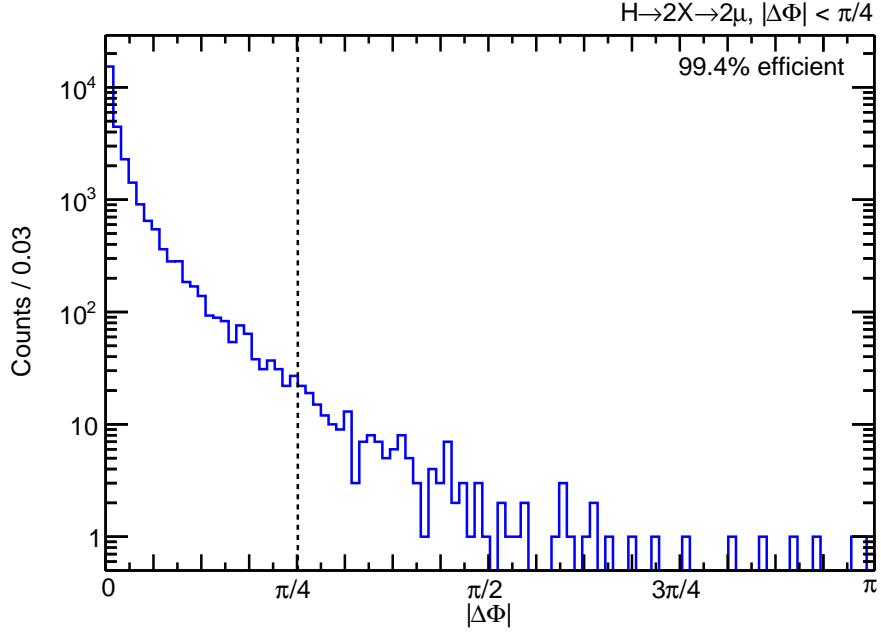


Figure 4.16: Histogram of  $|\Delta\Phi|$  for all  $2\mu$  signal samples combined, showing a strong peak at zero. The dashed line is at  $|\Delta\Phi| = \pi/4$ ; the cut  $|\Delta\Phi| < \pi/4$  is 99.4% efficient in signal.

To validate the approximate DY background symmetry around  $\pi/2$  (and better quantify the transfer factor), the analysis considers pairs of control regions that include the regions  $|\Delta\Phi| < \pi/4$  or  $|\Delta\Phi| > 3\pi/4$  but are formed by reversing other selections (discussed below).

#### 4.5.1 Definition of Control Regions

One set of DSA dimuon control events is obtained by reversal of the  $L_{xy}/\sigma_{L_{xy}}$  selection, so that events have small  $L_{xy}/\sigma_{L_{xy}}$  instead:  $L_{xy}/\sigma_{L_{xy}} < 6$ . Another set of DSA dimuon events is obtained by reversal of the opposite-sign charge selection, so that dimuons consist of muons reconstructed with the same sign.

Yet another set of DSA dimuon events is obtained by reversal of DSA to PAT association. This set consists of DSA dimuons (formed from two DSA muons), passing all selections except for the DSA to PAT association procedure, in which the two constituent DSA muons can be associated with two PAT muons, and the resulting dimuon formed from two PAT muons

has its own PAT  $L_{xy}/\sigma_{L_{xy}}$  satisfying a certain requirement.

Two such PAT  $L_{xy}/\sigma_{L_{xy}}$  requirements are used in this analysis. In the first set of events, the PAT  $L_{xy}/\sigma_{L_{xy}}$  is small, e.g.  $L_{xy}/\sigma_{L_{xy}} < 1$ . This region is designed to study Drell-Yan background, and corresponds to DSA dimuons whose PAT muon counterparts are consistent with being produced promptly (and thus are not signal), but which may be reconstructed as highly displaced due to reconstruction mistakes or vertex fit anomalies. In order to minimize contamination from QCD events, each constituent PAT muon is required to be isolated. The relative isolation of a muon track is defined as the scalar sum of the  $p_T$  of tracks in a cone around the muon divided by the muon  $p_T$ . For this first range, the relative isolation of each constituent PAT muon is required to be at most 0.05.

In the second set of events, the PAT  $L_{xy}/\sigma_{L_{xy}}$  is large, such as  $60 < L_{xy}/\sigma_{L_{xy}} < 115$ . This region is designed to study QCD background, such as from cascade decays of  $b$  to  $c$  quarks or hadrons in flight, with  $c\tau$  values of 0.5–2 millimeters and with typical PAT-PAT dimuon  $L_{xy}$  resolutions of 10–30  $\mu\text{m}$ . In order to minimize signal contamination, each constituent PAT muon is required *not* to be isolated, meaning the relative isolation of each constituent PAT muon is required to be at least 0.5. Less than 3% of  $2\mu$  signal events satisfy this non-isolation requirement, which meanwhile selects a subset of events composed of approximately 95% events from QCD processes (according to MC simulation).

With any combination of these three reversals, a pair of control regions may be formed satisfying  $|\Delta\Phi| < \pi/4$  or  $|\Delta\Phi| > 3\pi/4$ , respectively.

Some tedious notation is introduced to facilitate discussion of the various control regions. These regions are labeled CR with the following sublabels:

- The superscript on the left side of CR is either 0, 1, 2, or  $\pi$ , corresponding respectively to the  $|\Delta\Phi|$  ranges  $0 < |\Delta\Phi| < \pi/4$ ,  $\pi/4 < |\Delta\Phi| < \pi/2$ ,  $\pi/2 < |\Delta\Phi| < 3\pi/4$ , and  $3\pi/4 < |\Delta\Phi| < \pi$ .
- The subscript on the left side of CR is either OS or SS, corresponding respectively to DSA dimuons consisting of muons with charges of **o**pposite **s**ign and muons with charges of the **s**ame **s**ign.



- The superscript on the right side of CR is of the format  $L < x$  or  $L > x$  with  $x$  a cut value, usually 6, corresponding to the  $L_{xy}/\sigma_{L_{xy}}$  requirement for DSA dimuons. This label means that  $L_{xy}/\sigma_{L_{xy}}$  value for the dimuon formed from two DSA muons is less than  $x$  or greater than  $x$ . The  $L$  in this expression is a reminder that the cut value is for  $L_{xy}/\sigma_{L_{xy}}$ .
- The subscript on the right side of CR is one of
  - full, corresponding to a control region in which the full, standard DSA to PAT association is performed;
  - DY, corresponding to the DSA dimuons with reversed DSA to PAT association as described above with the PAT dimuon satisfying  $L_{xy}/\sigma_{L_{xy}} < 1$  and muons isolated; or
  - QCD, corresponding to the DSA dimuons with reversed DSA to PAT association as described above with the PAT dimuon satisfying  $60 < L_{xy}/\sigma_{L_{xy}} < 115$  and muons not isolated.

The DSA dimuon passes the selections on  $|\Delta\Phi|$  and  $L_{xy}/\sigma_{L_{xy}}$  given by the other two labels. The DY and QCD in these expressions indicate the type of background that these control regions are designed to help estimate; the PAT  $L_{xy}/\sigma_{L_{xy}}$  ranges were chosen accordingly.

For example,  $\pi_{\text{OS}}\text{CR}_{\text{DY}}^{L<6}$  corresponds to opposite-sign DSA dimuons with  $|\Delta\Phi| > 3\pi/4$ , DSA  $L_{xy}/\sigma_{L_{xy}} < 6$ , and reversed DSA to PAT association with PAT  $L_{xy}/\sigma_{L_{xy}} < 1$ . The signal region SR consisting of DSA dimuons passing all selections can be expressed in this notation, with SR replacing CR, as  $\text{SR} \equiv {}^0_{\text{OS}}\text{SR}_{\text{full}}^{L>6}$ .

#### 4.5.2 Estimation of Drell-Yan Background

Misreconstructed Drell-Yan events form a large majority of the background in most regions. Figure 4.17 shows distributions of  $|\Delta\Phi|$  in data and in MC simulation for the DY control regions with DSA to PAT association reversed, for all ranges of  $|\Delta\Phi|$ . The left plot is for

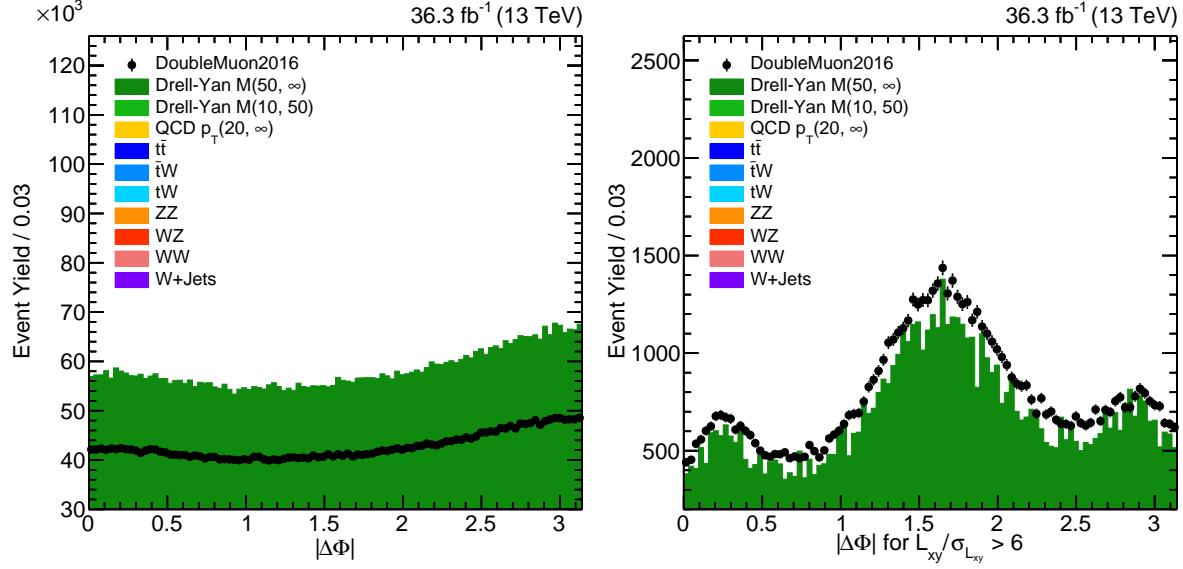


Figure 4.17: Histograms of  $|\Delta\Phi|$  in data and in MC simulation for the DY control regions with DSA to PAT association reversed, for all values of  $|\Delta\Phi|$ , for (left) all values of  $L_{xy}/\sigma_{L_{xy}}$  and (right) only  $L_{xy}/\sigma_{L_{xy}} > 6$ . In the previously introduced notation, (right) consists of  ${}^0_{\text{OS}}\text{CR}_{\text{DY}}^{L>6}$ ,  $\pi_{\text{OS}}\text{CR}_{\text{DY}}^{L>6}$ , etc., and (left) consists of all the events in the right plot plus  ${}^0_{\text{OS}}\text{CR}_{\text{DY}}^{L<6}$ ,  $\pi_{\text{OS}}\text{CR}_{\text{DY}}^{L<6}$ , etc. Without any  $L_{xy}/\sigma_{L_{xy}}$  cut, the distribution is somewhat symmetric, with some excess of events at  $\pi$  compared to 0. With the  $L_{xy}/\sigma_{L_{xy}}$  cut, events near 0,  $\pi/2$ , and  $\pi$  are suppressed less than other events, although the distributions are still fairly symmetric.

all values of  $L_{xy}/\sigma_{L_{xy}}$ ; that is, it contains both  ${}^0_{\text{OS}}\text{CR}_{\text{DY}}^{L>6}$  and  ${}^0_{\text{OS}}\text{CR}_{\text{DY}}^{L<6}$  for all  $|\Delta\Phi|$  ranges (0, 1, 2,  $\pi$ ). The right plot is only for  $L_{xy}/\sigma_{L_{xy}} > 6$ ; that is, it only contains  ${}^0_{\text{OS}}\text{CR}_{\text{DY}}^{L>6}$  for all  $|\Delta\Phi|$  ranges. Without any  $L_{xy}/\sigma_{L_{xy}}$  cut, the distribution is somewhat symmetric, with some excess of events at  $\pi$  compared to 0. With the  $L_{xy}/\sigma_{L_{xy}}$  cut, events at 0,  $\pi/2$ , and  $\pi$  are suppressed less than other events, although the distributions are still fairly symmetric.

Understanding this shape begins by considering Drell-Yan events, which compose a majority of the background, and for which the event geometry often consists of two back-to-back muons. These muons' directions are often well reconstructed, whereas their  $p_T$  may not be, resulting in the dimuon  $p_T$  pointing along the direction of one of the muons. With this event topology in mind, the shape can be explained by the combination of two effects.

The first effect is a result of the decrease in trigger efficiency for large transverse impact parameters  $d_0$ . Events in which the muon tracks are parallel to the  $L_{xy}$  vector are events in which the muons point towards the primary vertex, so that their  $d_0$  is small. Events in which the muon tracks are perpendicular to the  $L_{xy}$  vector are events with large transverse impact parameters, and would be suppressed by the trigger. Since the dimuon  $p_T$  is typically along the direction of one of the muons, muon tracks parallel (perpendicular) to the  $L_{xy}$  vector correspond to  $|\Delta\Phi| \approx 0$  or  $\pi$  ( $|\Delta\Phi| \approx \pi/2$ ). In other words, events with large values of  $L_{xy}$  only pass the trigger when the muons sufficiently point towards the primary vertex, i.e. have  $|\Delta\Phi| \approx 0$  (or  $\pi$ ), effectively suppressing  $|\Delta\Phi| \approx \pi/2$ .

(As a thought experiment, consider muons moving in the detector with no magnetic field. Then  $d_0 = L_{xy} \sin |\Delta\Phi|$ . If the trigger efficiency sharply vanished at some fixed value  $d_0 \geq D$ , then only events with  $L_{xy} \leq D \csc |\Delta\Phi|$  would pass the trigger. Cosecant takes large values around  $|\Delta\Phi| = 0$  and  $|\Delta\Phi| = \pi$ , and small values at  $|\Delta\Phi| = \pi/2$ . The diagram in the top row of Figure 4.18 illustrates this first effect in the context of this thought experiment.)

The second effect is a result of geometric effects on the uncertainty of the position of the fitted vertex. For back-to-back muons, the uncertainty on the fitted dimuon vertex is larger along the direction of the tracks than orthogonal to it, because the uncertainty orthogonal to the track is set by the scatter of hits about the tracks, whereas determining the position of the vertex along two back-to-back tracks is more difficult. Then for a given  $L_{xy}$ , the uncertainty on  $L_{xy}$ ,  $\sigma_{L_{xy}}$ , is larger when the muons (and hence the dimuon  $p_T$ ) are parallel to the  $L_{xy}$  vector (i.e.  $|\Delta\Phi| = 0$  or  $\pi$ ) and smaller when the muons (and hence the dimuon  $p_T$ ) are perpendicular to the  $L_{xy}$  vector (i.e.  $|\Delta\Phi| = \pi/2$ ). In other words,  $\sigma_{L_{xy}}$  is smaller at  $\pi/2$  than at 0 or  $\pi$ . The diagrams in the bottom row of Figure 4.18 illustrate this second effect.

A selection on  $L_{xy}/\sigma_{L_{xy}}$  selects events with large  $L_{xy}$  and/or small  $\sigma_{L_{xy}}$ , which have been shown to preferentially occur at 0,  $\pi/2$ , and  $\pi$ . The resulting shape of the  $|\Delta\Phi|$  distribution appears to be geometric in origin, and is the motivation for defining the signal and control regions as  $|\Delta\Phi| < \pi/4$  and its symmetric region around  $\pi/2$ , in order to cut away from the extra background at  $\pi/2$ . Figure 4.19 shows the  $|\Delta\Phi|$  distributions as in the right plot of

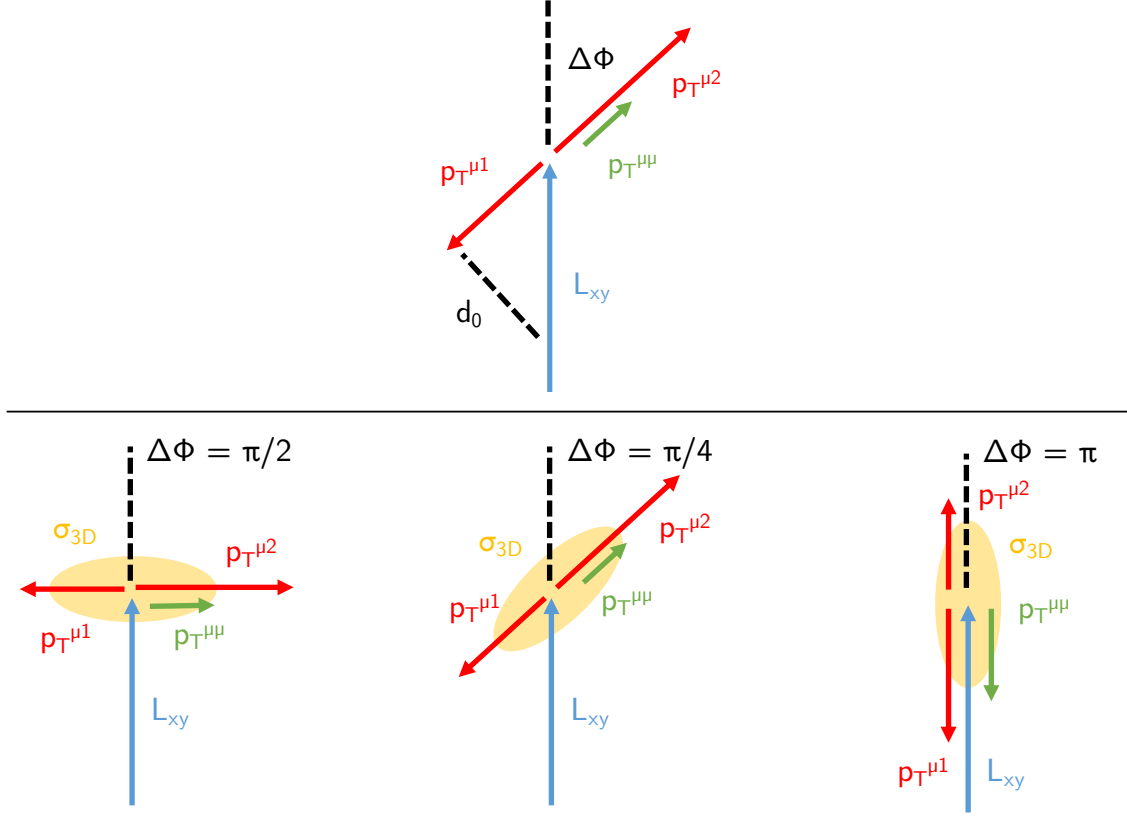


Figure 4.18: (Top) Diagram depicting the effect of dimuon orientation on possible values of  $L_{xy}$ . Lowered trigger efficiency at large transverse impact parameter ( $d_0$ ) values suppresses large  $L_{xy}$  values, which have significant  $d_0$  unless  $|\Delta\Phi|$  is sufficiently close to 0 or  $\pi$ . (Bottom) Diagram depicting the effect of dimuon orientation on  $\sigma_{L_{xy}}$  as a function of  $|\Delta\Phi|$ . The dimuon  $p_T$  points along one of the muons, which are back-to-back, and the uncertainty on the position of the fitted vertex is represented by a yellow ellipse around the vertex, with the larger axis parallel to the muon. For a given  $L_{xy}$  the  $\sigma_{L_{xy}}$  is larger when the muons are parallel to the  $L_{xy}$  vector (i.e.  $|\Delta\Phi| = 0$  or  $\pi$ ) and smaller when the muons are perpendicular to it (i.e.  $|\Delta\Phi| = \pi/2$ ), taking an intermediate value with  $|\Delta\Phi| = \pi/4$  (or  $3\pi/4$ ).

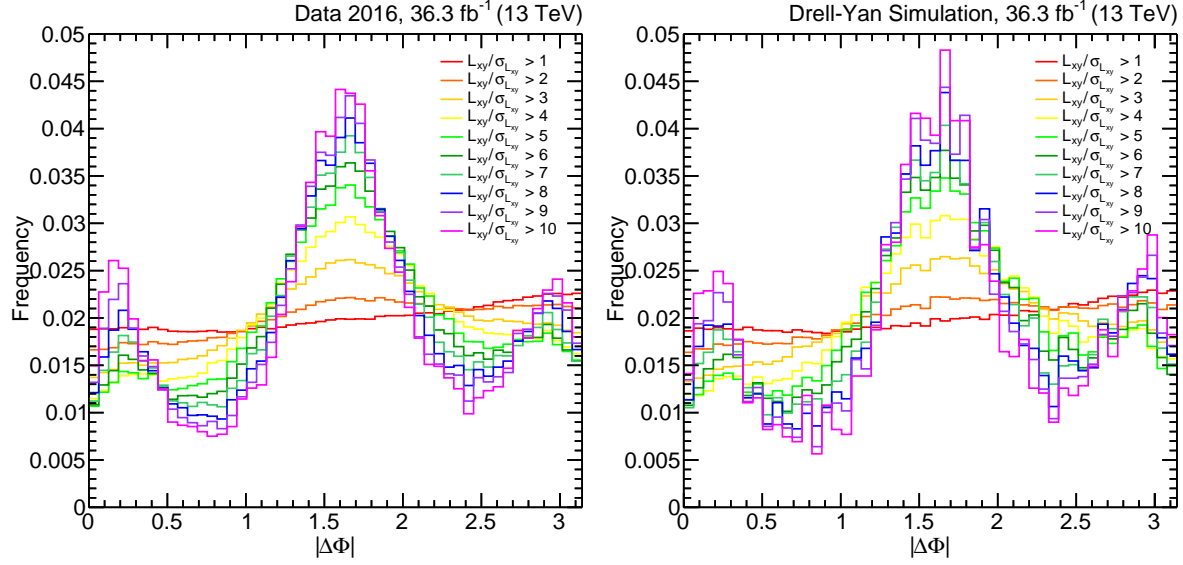


Figure 4.19: Histograms of  $|\Delta\Phi|$  as in Figure 4.17, normalized to unit area, for sequentially increasing cut values of  $L_{xy}/\sigma_{L_{xy}}$  for (left) data and (right) Drell-Yan simulation.

Figure 4.17, each scaled to unit area, for data and for Drell-Yan simulation, for sequentially increasing cuts on  $L_{xy}/\sigma_{L_{xy}}$  (i.e.  ${}^0\text{OSCR}_{\text{DY}}^{L>x}$  for all  $|\Delta\Phi|$  and  $x$  variable). The Drell-Yan plot indicates that the shape is well-reproduced by simulation.

Table 4.5 gives the number of events in data for each of the control regions. A data-driven estimate of the transfer factor  $\tau_{\text{DY}}$  can be obtained by dividing the number of events in  ${}^0\text{OSCR}_{\text{DY}}^{L>6}$  by the number of events in  ${}^\pi\text{OSCR}_{\text{DY}}^{L>6}$  and is found to be

$$\tau_{\text{DY}} = \frac{N[{}^0\text{OSCR}_{\text{DY}}^{L>6}]}{N[{}^\pi\text{OSCR}_{\text{DY}}^{L>6}]} = 0.78 \quad (4.6)$$

Other ratios, such as  ${}^0\text{OSCR}_{\text{DY}}^{L<6}/{}^\pi\text{OSCR}_{\text{DY}}^{L<6} = 0.89$  and  ${}^0\text{OSCR}_{\text{full}}^{L<6}/{}^\pi\text{OSCR}_{\text{full}}^{L<6} = 1.09$  serve to validate the (approximate) symmetry in the Drell-Yan background (because the values are close to 1).

There are 0 events in the full 2016 dataset in  ${}^\pi\text{OSCR}_{\text{full}}^{L>6}$ . The estimate of the expected Drell-Yan background in SR is therefore also 0 events (with or without the mass window cuts

Name	Events
$\pi_{\text{OS}}\text{CR}_{\text{full}}^{L>6}$	0
$^0_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$	60
$\pi_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$	55
$^0_{\text{OS}}\text{CR}_{\text{DY}}^{L>6}$	13636
$\pi_{\text{OS}}\text{CR}_{\text{DY}}^{L>6}$	17405
$^0_{\text{OS}}\text{CR}_{\text{DY}}^{L<6}$	1026522
$\pi_{\text{OS}}\text{CR}_{\text{DY}}^{L<6}$	1153133

Table 4.5: Event counts for control regions for estimating Drell-Yan background. The ratio of  $^0_{\text{OS}}\text{CR}_{\text{DY}}^{L>6}$  to  $\pi_{\text{OS}}\text{CR}_{\text{DY}}^{L>6}$  gives a transfer factor; multiplying the number of events in  $\pi_{\text{OS}}\text{CR}_{\text{full}}^{L>6}$  by this transfer factor gives an estimate of the Drell-Yan background in SR.

described in Section 4.6.2). The systematic uncertainty on  $\tau_{\text{DY}}$  is described and evaluated in Section 4.7.2.

### 4.5.3 Estimation of QCD Background

Background events arising from QCD processes, such as from collimated muons in jets, hadron decays in flight, or cascade decays of  $b$  and  $c$  quarks, may have large, signal-like  $L_{\text{xy}}/\sigma_{L_{\text{xy}}}$  values and low dimuon masses near 10–20 GeV.  $|\Delta\Phi|$  for such events may not be symmetric around  $\pi/2$ , instead having a peak at 0.

Figure 4.20 illustrates this with a histogram of  $|\Delta\Phi|$  for opposite sign events with the DSA to PAT association reversed, with the corresponding matched PAT dimuon satisfying  $60 < L_{\text{xy}}/\sigma_{L_{\text{xy}}} < 115$ , and the constituent PAT muon tracks not isolated, for sequentially increasing cuts on DSA  $L_{\text{xy}}/\sigma_{L_{\text{xy}}}$ . Unlike the DY-like events in the data plot of Figure 4.19, which are symmetric around  $\pi/2$ , these events have a strong, signal-like peak in  $|\Delta\Phi|$  at 0, especially for larger values of  $L_{\text{xy}}/\sigma_{L_{\text{xy}}}$ .

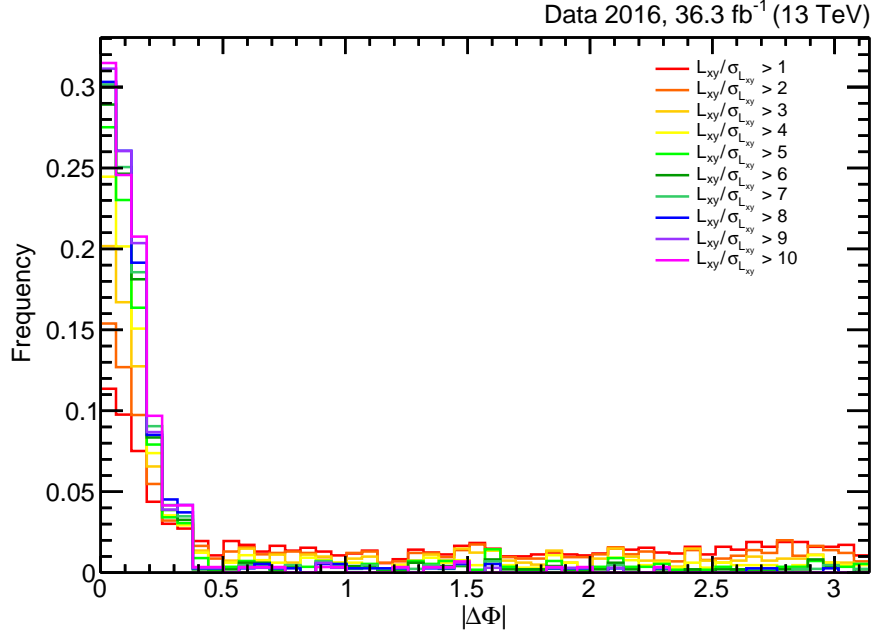


Figure 4.20: Histogram of  $|\Delta\Phi|$  for sequentially increasing cut values of  $L_{xy}/\sigma_{L_{xy}}$  in data for the control region  ${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L>x}$ , normalized to unit area, with variable, sequentially increasing cut values  $x$  of DSA  $L_{xy}/\sigma_{L_{xy}}$ . These events are QCD-like, and have a strong, signal-like peak in  $|\Delta\Phi|$  at 0, especially for larger values of  $L_{xy}/\sigma_{L_{xy}}$ , in stark contrast to the symmetry around  $\pi/2$  of the DY-like events in the data plot of Figure 4.19.

To demonstrate that describing the two sets of events (with small and large PAT  $L_{xy}/\sigma_{L_{xy}}$  requirements) as DY-like and QCD-like is justified, Figure 4.21 shows histograms of  $m_{\mu\mu}$  and  $\Delta R(\mu\mu)$  for sequentially increasing cut values of  $L_{xy}/\sigma_{L_{xy}}$  in data, as in Figure 4.19, for the events in the control region (i.e.  ${}^0_{\text{OS}}\text{CR}_{\text{DY}}^{L>x}$  for variable  $x$ ) used to estimate the Drell-Yan background in Section 4.5.2. The dimuon mass plot features a Z mass peak, and the  $\Delta R$  plot is highly DY-like, with a strong peak at  $\pi$ . As the cut value increases, the contributions of the Z mass peak and the  $\Delta R$  peak at  $\pi$  decrease, and the contributions of QCD events at lower mass and small  $\Delta R$  increase slightly.

In contrast, Figure 4.22 shows similar histograms of  $m_{\mu\mu}$  and  $\Delta R(\mu\mu)$  for the QCD-like events in data (i.e.  ${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L>x}$  for variable  $x$ ). As the cut increases, the contribution of lower mass QCD events increases, and the  $\Delta R$  distribution has a strong peak at 0, suggestive of QCD background, with successively decreasing contributions from Drell-Yan backgrounds at  $\pi$ .

Such QCD-like events hence would not appear in control regions with low  $L_{xy}/\sigma_{L_{xy}}$  or high  $|\Delta\Phi|$ , and estimation of this background cannot be performed with a transfer factor applied to events in  ${}^\pi_{\text{OS}}\text{CR}_{\text{full}}^{L>6}$ .

Instead, pairs of regions with same-sign dimuons and opposite-sign dimuons are considered. Table 4.6 gives the number of events in data for several control regions. The relevant control regions are then  ${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L>6}$  and  ${}^0_{\text{SS}}\text{CR}_{\text{QCD}}^{L>6}$ , and the transfer factor thus obtained can be applied to  ${}^0_{\text{SS}}\text{CR}_{\text{full}}^{L>6}$  to obtain a QCD background estimate for  ${}^0_{\text{OS}}\text{CR}_{\text{full}}^{L>6}$ . This transfer factor has a value of

$$\tau_{\text{QCD}} = \frac{N[{}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L>6}]}{N[{}^0_{\text{SS}}\text{CR}_{\text{QCD}}^{L>6}]} = 2.81 \quad (4.7)$$

There are 4 same-sign events in  ${}^0_{\text{SS}}\text{CR}_{\text{QCD}}^{L>6}$  in the full 2016 dataset, giving an estimate of QCD background in the SR of 11.2 events (without the mass window cuts described in Section 4.6.2). As with the Drell-Yan transfer factor, there is a large systematic uncertainty on the QCD transfer factor (among other factors, it is sensitive to the window used for the PAT  $L_{xy}/\sigma_{L_{xy}}$ ); this uncertainty is described and evaluated in Section 4.7.2.



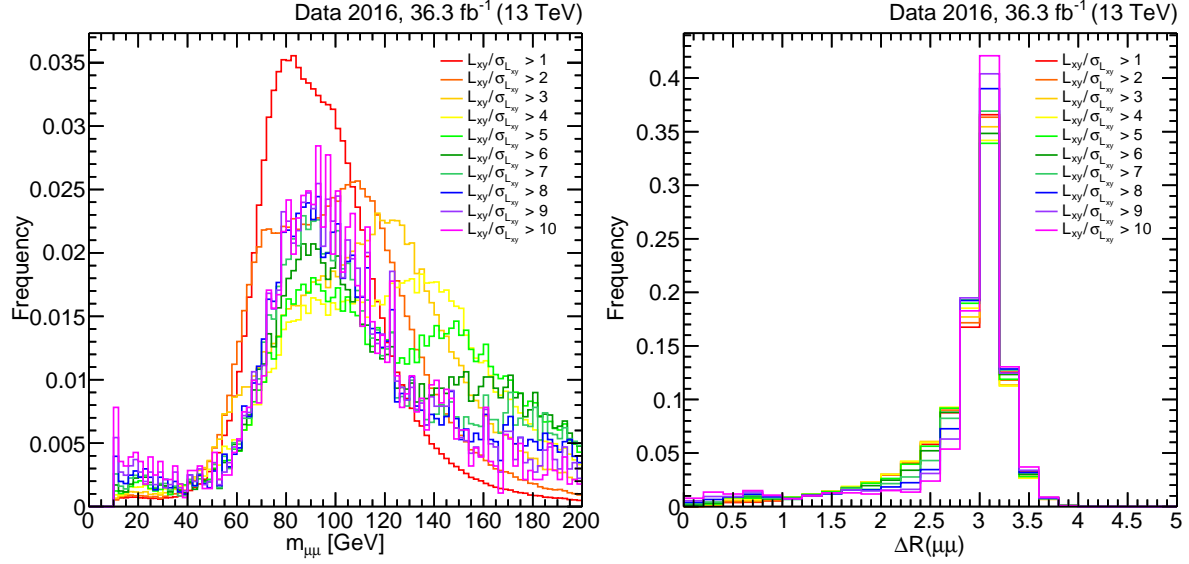


Figure 4.21: Histograms of (left)  $m_{\mu\mu}$  and (right)  $\Delta R(\mu\mu)$  for sequentially increasing cut values of  $L_{xy}/\sigma_{L_{xy}}$  in data as in Figure 4.19, normalized to unit area, using the control regions  ${}^0_{\text{OS}}\text{CR}_{\text{DY}}^{L>x}$  for variable, sequentially increasing cut values  $x$  of DSA  $L_{xy}/\sigma_{L_{xy}}$ .

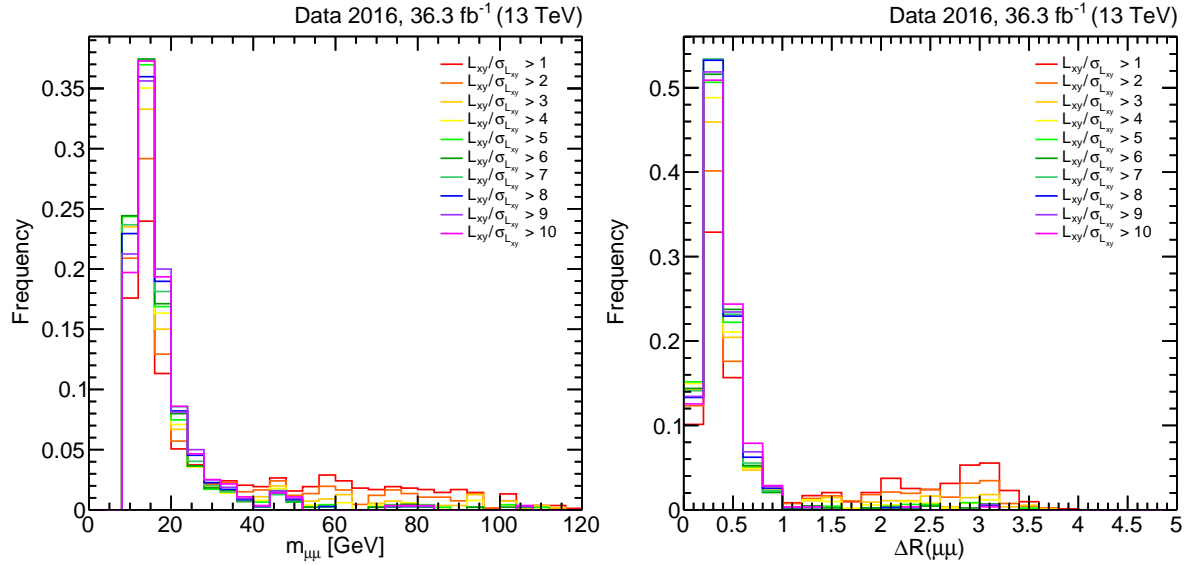


Figure 4.22: Histograms of (left)  $m_{\mu\mu}$  and (right)  $\Delta R(\mu\mu)$  as in Figure 4.21, but for QCD-like events in data, using the control regions  ${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L>x}$  for variable, sequentially increasing cut values  $x$  of DSA  $L_{xy}/\sigma_{L_{xy}}$ .

Name	Events
${}^0_{\text{SS}}\text{CR}_{\text{full}}^{L>6}$	4
${}^0_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$	60
${}^0_{\text{SS}}\text{CR}_{\text{full}}^{L<6}$	4
${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L>6}$	323
${}^0_{\text{SS}}\text{CR}_{\text{QCD}}^{L>6}$	115
${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L<6}$	314
${}^0_{\text{SS}}\text{CR}_{\text{QCD}}^{L<6}$	196

Table 4.6: Event counts for control regions for estimating QCD background. A ratio of events in  ${}^0_{\text{SS}}\text{CR}_{\text{QCD}}^{L>6}$  with same-sign muons to the corresponding control region with opposite-sign muons  ${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L>6}$  gives a transfer factor; multiplying the number of events in  ${}^0_{\text{SS}}\text{CR}_{\text{full}}^{L>6}$  by this transfer factor gives an estimate of the QCD background in SR.

Other ratios in Table 4.5 do not give an appropriate transfer factor but rather serve to illuminate features of the QCD background. The ratio of  ${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L<6}$  to  ${}^0_{\text{SS}}\text{CR}_{\text{QCD}}^{L<6}$  is 1.6, much closer to 1; this clarifies that there is proportionally more opposite-sign QCD background for large values of  $L_{xy}/\sigma_{L_{xy}}$ , and so a large  $L_{xy}/\sigma_{L_{xy}}$  region needed to be used to quantify it. The ratio of  ${}^0_{\text{SS}}\text{CR}_{\text{full}}^{L<6}$  to  ${}^0_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$  is quite small, showing that the QCD background is not an important contributor to the regions with small  $L_{xy}/\sigma_{L_{xy}}$ .

## 4.6 Cut Optimization

This section describes a few procedures applied to fine-tune the event and object selection to optimize the analysis for the statistical significance of a potential signal discovery.

#### 4.6.1 Optimizing with $Z_{\text{Bi}}$ as a Figure of Merit

As a set of analysis selections varies, the numbers of signal and background events also vary, and so fine-tuning the values of cuts is an exercise in balancing signal efficiency vs. background rejection. The metric for determining an optimal cut value, as a function of the number of signal and background events, is a figure of merit that is monotonic with the expected statistical significance of a discovery. The figure of merit used in this analysis is  $Z_{\text{Bi}}$  [53], an estimate of the statistical significance of an excess number of events in a signal region when the number of background events is estimated from a control region. The subscript is an abbreviation for **B**inomial, because under the background-only hypothesis, the splitting of events between the signal and control region follows a binomial distribution with binomial parameter equal to the expected fraction of background in the signal region out of the background in both regions. In the notation introduced in Section 4.5, with  $\tau$  the transfer factor, this binomial parameter is  $1/(1 + \tau)$ .

A procedure based on  $Z_{\text{Bi}}$  as a figure of merit was performed on several key discriminating variables. The following description of the procedure is phrased in terms of optimizing the  $L_{xy}/\sigma_{L_{xy}}$  cut. A background distribution of  $L_{xy}/\sigma_{L_{xy}}$  is defined as a histogram of data events in the (approximately) signal-free region  $|\Delta\Phi| > \pi/2$ , and a corresponding signal distribution of  $L_{xy}/\sigma_{L_{xy}}$  is defined as a histogram of signal events in the signal region  $|\Delta\Phi| < \pi/2$ , with both distributions consisting of events passing all selections except for the  $L_{xy}/\sigma_{L_{xy}}$  cut. The signal histogram is scaled to 2016 integrated luminosity assuming an arbitrary production cross section of  $1 \times 10^{-2}$  pb. For  $L_{xy}/\sigma_{L_{xy}}$ , the number of events in either histogram passing a given cut is an integral from the cut value to infinity; for some other variables (such as track  $\chi^2/\text{dof}$ ), the number of events passing a given cut is an integral from zero to the cut value. In the following discussion, the “integral” of a histogram is assumed to be performed over the appropriate interval.

In the notation found in Reference [53], the integral of the simulated signal histogram corresponds to  $\mu_s$ , the true Poisson mean number of signal events in the signal region; and the integral of the background histogram corresponds to a number of observed events in the

signal-free region  $n_{\text{off}}$ , from which one derives the estimate  $\hat{\mu}_b = n_{\text{off}}/\tau$  of the true Poisson mean number of background events in the signal region. To obtain the median  $Z_{\text{Bi}}$ ,  $\mu_s$  is used as the estimate of the true Poisson signal mean  $\hat{\mu}_s$ .  $Z_{\text{Bi}}$  is then calculated for each cut value, a function of three quantities: the number of background events in the control region  $n_{\text{off}}$ , the total number of observed events in the signal region  $n_{\text{on}} = \hat{\mu}_s + \hat{\mu}_b$ , and the transfer factor is taken to be  $\tau = 1$ .

Figure 4.23 shows an example of the cut optimization procedure, with the background (control region in data) distribution in red, the signal distribution (here for the  $2\mu$  signal sample with  $m_{\text{H}} = 1000 \text{ GeV}$ ,  $m_{\text{X}} = 20 \text{ GeV}$ , and  $c\tau = 20 \text{ mm}$ ) in blue, and the value of  $Z_{\text{Bi}}$  for each cut value in green. The tail of the signal distribution is much longer than in the background distribution, which peaks strongly at smaller values of  $L_{\text{xy}}/\sigma_{L_{\text{xy}}}$ , so  $Z_{\text{Bi}}$  rises steadily until about 5 or 6. For this signal sample, the cut value of  $L_{\text{xy}}/\sigma_{L_{\text{xy}}}$  corresponding to the largest value of  $Z_{\text{Bi}}$  is between 6 and 7.

This procedure was performed on all signal samples, with a systematic grid of cut values in several variables. The main results of this  $Z_{\text{Bi}}$ -based cut optimization are:

- $Z_{\text{Bi}}$  rises steadily as cuts on  $L_{\text{xy}}/\sigma_{L_{\text{xy}}}$  and track  $\chi^2/\text{dof}$  are tightened, informing the choices of cut values of 6 and 2.5, respectively. Although for some sets of selections,  $Z_{\text{Bi}}$  continues to increase for tighter cuts, the shape flattens out considerably after these thresholds, so the increases in  $Z_{\text{Bi}}$  are minimal, and the shape becomes highly sensitive to the specific background events and the specific numbers of events used to tune the cuts. The choices of cut values take these effects into account.
- $Z_{\text{Bi}}$  does not increase appreciably for any requirement on the muon track impact parameter significance ( $d_0/\sigma_{d_0}$ ) once a requirement on  $L_{\text{xy}}/\sigma_{L_{\text{xy}}}$  requirement is in place, so the analysis selection does not impose any cuts on  $d_0/\sigma_{d_0}$ . This was a requirement that was used in previous versions of this analysis performed with Run 1 data [1, 2]. Similarly, no appreciable increase in  $Z_{\text{Bi}}$  was observed for tighter cuts on vertex  $\chi^2$  than 20 once other cuts are in place.

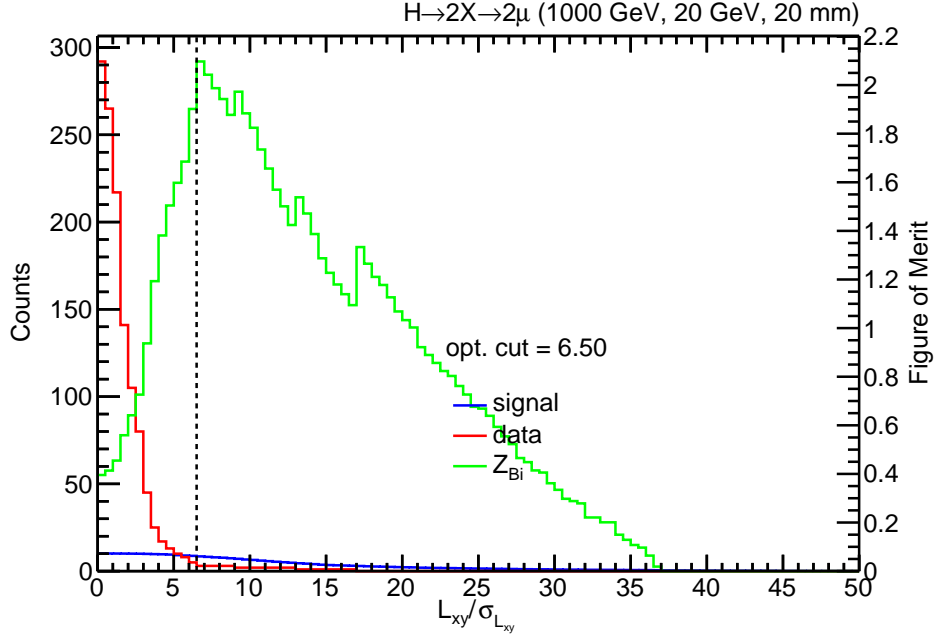


Figure 4.23: Example of  $Z_{\text{Bi}}$ -based cut optimization of  $L_{xy}/\sigma_{L_{xy}}$  for the  $2\mu$  signal sample with  $m_H = 1000$  GeV,  $m_X = 20$  GeV, and  $c\tau = 20$  mm. The background distribution is in red and corresponds to a selection of events in a control region of data; the signal distribution is in blue; and the value of  $Z_{\text{Bi}}$  at each cut value is in green. Here, the cut value of  $L_{xy}/\sigma_{L_{xy}}$  corresponding to the largest value of  $Z_{\text{Bi}}$  is between 6 and 7, marked on the plot as 6.5. As a consequence of the background distribution having a strong peak at small values of  $L_{xy}/\sigma_{L_{xy}}$  and the tail in the signal distribution being much longer, the  $Z_{\text{Bi}}$  curve rises steadily for tighter cuts on  $L_{xy}/\sigma_{L_{xy}}$ , peaks, and then falls.

Generated $m_X$	$\mu$	$\sigma$	Dimuon Mass Window
20 GeV	20 GeV	4 GeV	$10 \text{ GeV} < m_{\mu\mu} < 32 \text{ GeV}$
50 GeV	50 GeV	10 GeV	$20 \text{ GeV} < m_{\mu\mu} < 80 \text{ GeV}$
150 GeV	140 GeV	35 GeV	$35 \text{ GeV} < m_{\mu\mu} < 245 \text{ GeV}$
350 GeV	320 GeV	85 GeV	$65 \text{ GeV} < m_{\mu\mu}$

Table 4.7: Dimuon invariant mass window selections for each value of generated long-lived particle mass  $m_X$ , along with values for the fitted Gaussian  $\mu$  and  $\sigma$ . The windows are defined as  $\mu \pm 3\sigma$  with the lower limit at least 10 GeV and no upper limit for the largest generated mass.

#### 4.6.2 Dimuon Mass Window Selection

When computing upper limits on long-lived particle production cross sections, it is undesirable to include in the observation dimuon events with invariant masses appreciably different from the invariant mass postulated by the signal model under consideration. To implement this requirement, the reconstructed dimuon invariant mass histograms (for each value of generated long-lived particle mass  $m_X$ , for events passing the full selection in  $2\mu$  signal) are fit to Gaussian distributions, yielding a mean and standard deviation,  $\mu_{m_X}$  and  $\sigma_{m_X}$ . Figure 4.24 shows the distributions of reconstructed dimuon mass for  $2\mu$  signal samples, combining all sets of signal parameters for each value of  $m_X$ , for events passing all selections except for the mass cut, along with a fitted Gaussian for each distribution. The dimuon invariant mass is required to lie within a mass window given by  $\mu_{m_X} \pm 3\sigma_{m_X}$ , with the lower limit at least 10 GeV and no upper limit for the largest generated mass. This selection is more than 99% efficient in most  $2\mu$  signal samples (97% in the worst case). Table 4.7 enumerates the invariant mass window selection (applied to reconstructed dimuons) used for each generated value of  $m_X$ .

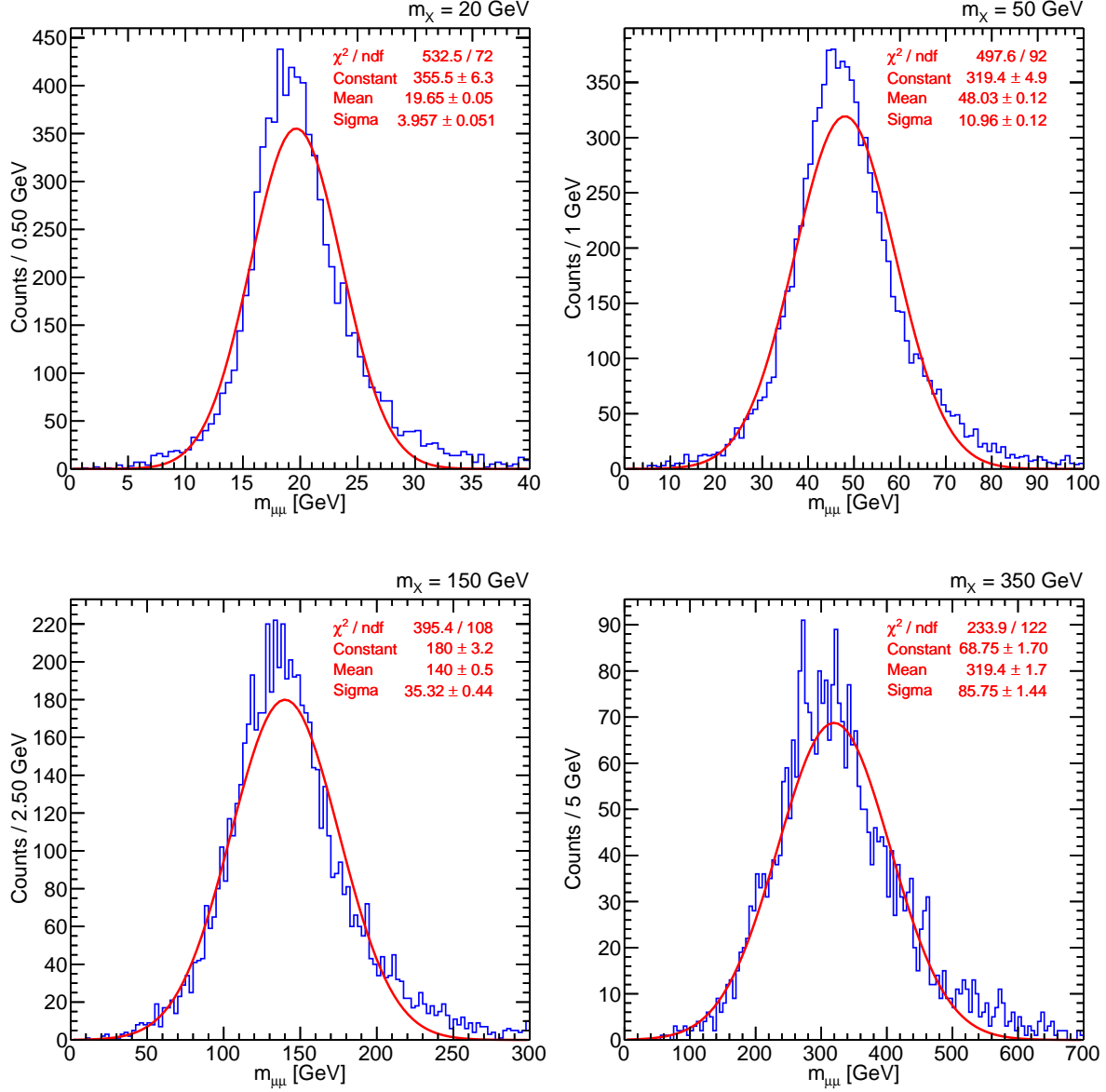


Figure 4.24: Reconstructed invariant dimuon mass distributions for  $2\mu$  signal samples passing all selections except the mass cut, combining all sets of signal parameters for each value of  $m_X$ , along with fitted Gaussian curves for each distribution.

### 4.6.3 N–1 Plots

This section presents selected “N–1” plots, so named because they are histograms of events passing the full selection except for one cut: the variable plotted. Such plots depict the events that passed all other cuts and would be removed by applying the cut in question, and so convey the effect of individual cuts in the analysis. These plots also provide a structured way of comparing the distributions and the fraction of events passing each cut in signal and background. The analysis selections should be highly efficient in signal and highly inefficient in background.

Figures 4.25–4.32 are the N–1 plots for selected variables, for events in all  $2\mu$  signal samples combined and for events in the control region  $\bar{\pi}_S \text{CR}_{\text{full}}^{L>6}$  of data. The cut values are labeled on the plots, as is the efficiency of the selection. The cosmic rejection cuts on  $\cos\alpha$  and number of parallel pairs are highly correlated, and so the corresponding plots, Figure 4.29 and Figure 4.30, omit both cuts from both sets of plots; for this reason they are actually “N–2” plots.

## 4.7 Systematic Uncertainties

This analysis produces an estimated number of signal events, given a signal model, and an estimated number of background events, given the background model. However, there are uncertainties associated with these estimates arising from a variety of sources; these uncertainties are known as systematic uncertainties, though most are ultimately statistical in origin. There are three main classes of systematic uncertainties in this analysis: the uncertainty on the measurement of the integrated luminosity, uncertainties associated with the signal estimate, and uncertainties associated with the background estimate.



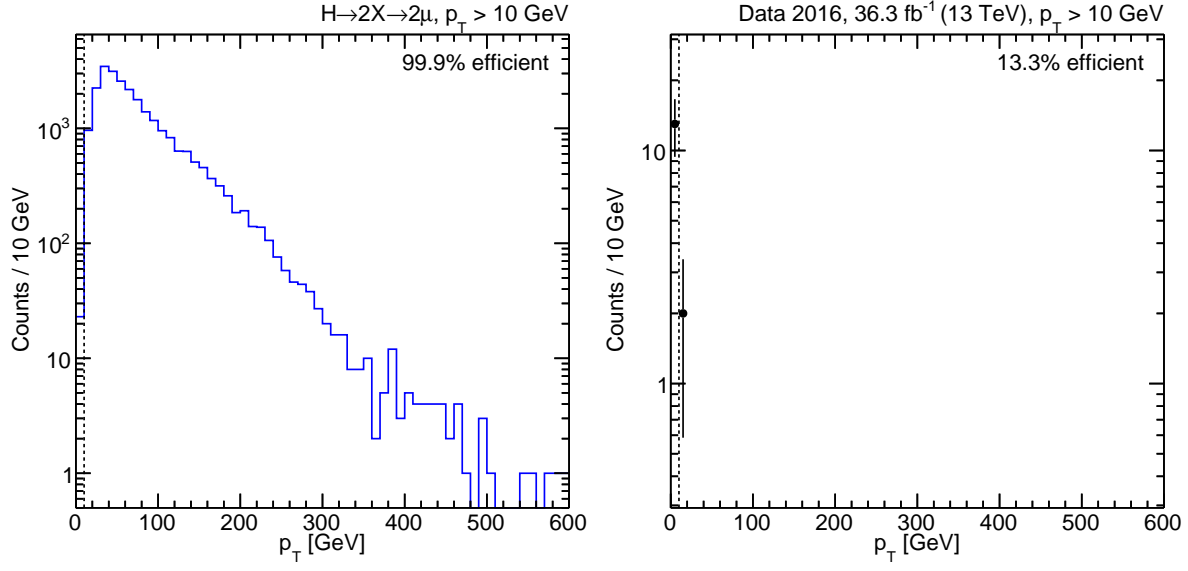


Figure 4.25: Histograms of events passing the full selection except for the muon  $p_T$  cut, in (left)  $2\mu$  signal and (right) data in the control region, along with the labeled cut value (and corresponding dashed line) and the selection efficiency.

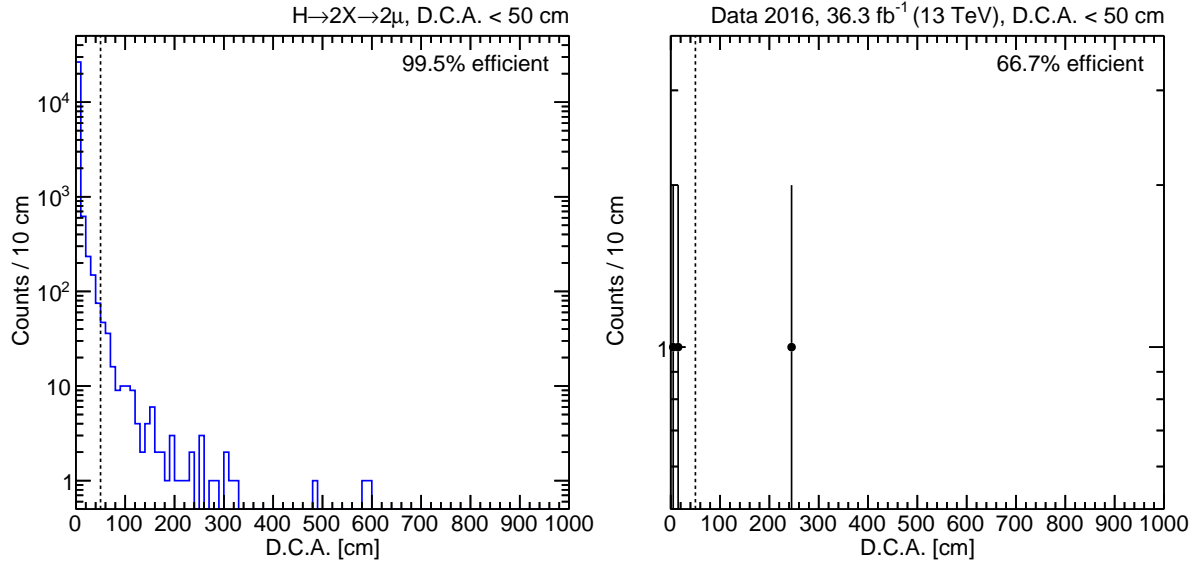


Figure 4.26: Histograms of events passing the full selection except for the distance of closest approach cut, in (left)  $2\mu$  signal and (right) data in the control region, along with the labeled cut value (and corresponding dashed line) and the selection efficiency.

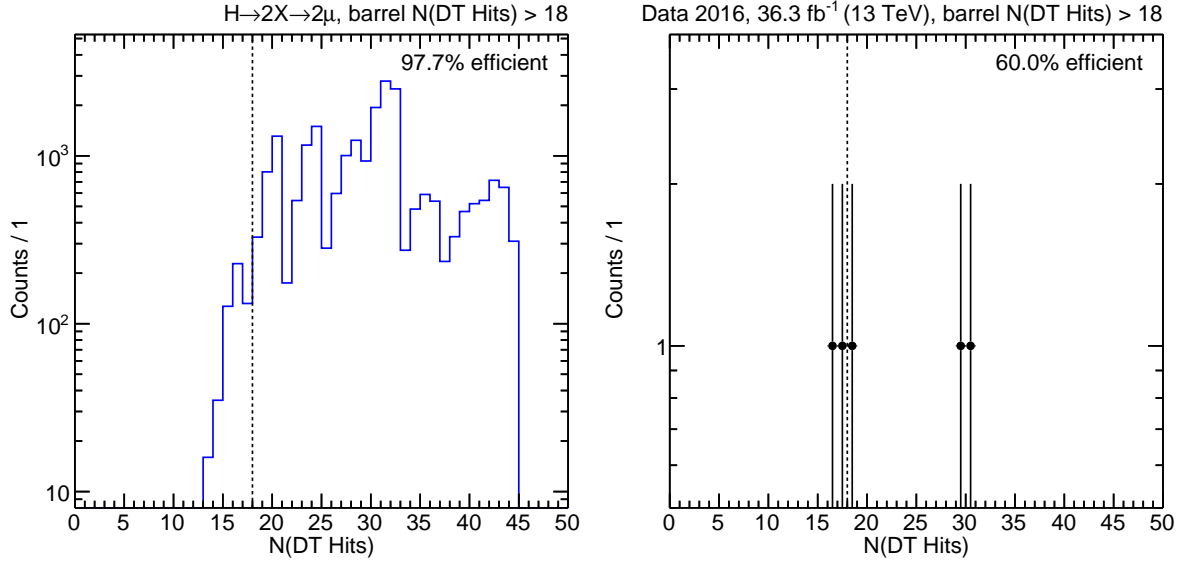


Figure 4.27: Histograms of events passing the full selection except for the barrel  $N(\text{DT hits})$  cut, in (left)  $2\mu$  signal and (right) data in the control region, along with the labeled cut value (and corresponding dashed line) and the selection efficiency.

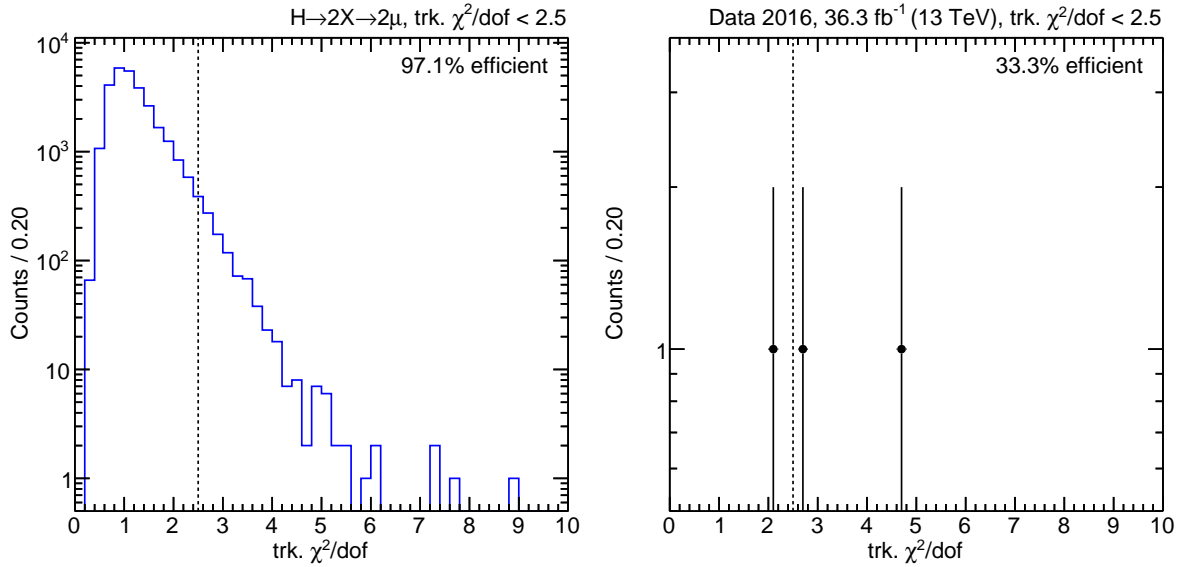


Figure 4.28: Histograms of events passing the full selection except for the track  $\chi^2/\text{dof}$  cut, in (left)  $2\mu$  signal and (right) data in the control region, along with the labeled cut value (and corresponding dashed line) and the selection efficiency.

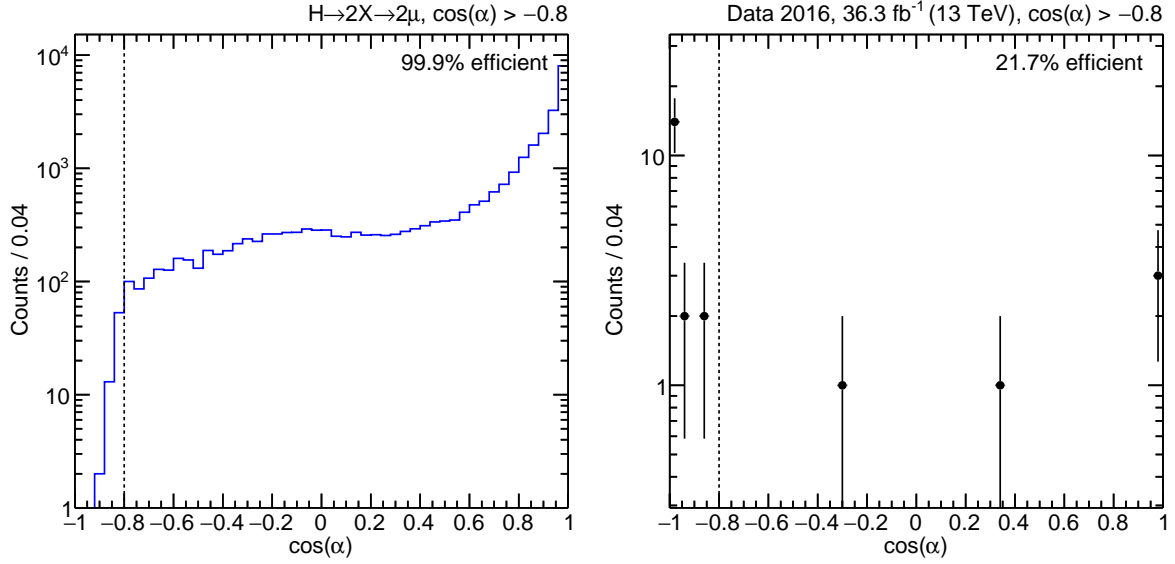


Figure 4.29: Histograms of events passing the full selection except for the  $\cos \alpha$  and  $N(\text{parallel pairs})$  cuts, in (left)  $2\mu$  signal and (right) data in the control region, along with the labeled cut value (and corresponding dashed line) and the selection efficiency.

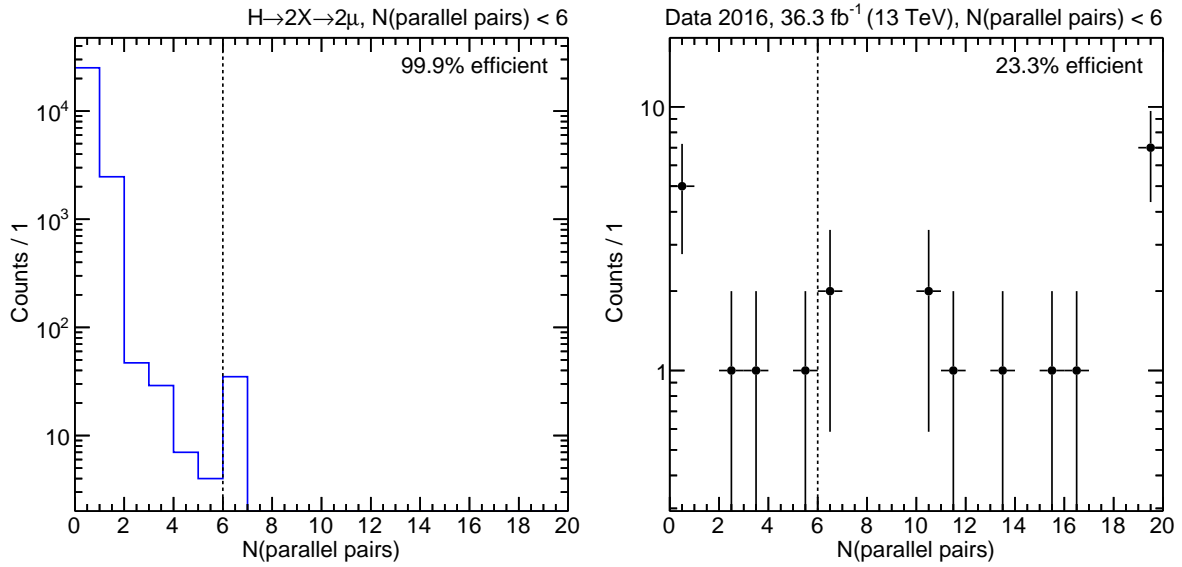


Figure 4.30: Histograms of events passing the full selection except for the  $\cos \alpha$  and  $N(\text{parallel pairs})$  cuts, in (left)  $2\mu$  signal and (right) data in the control region, along with the labeled cut value (and corresponding dashed line) and the selection efficiency.

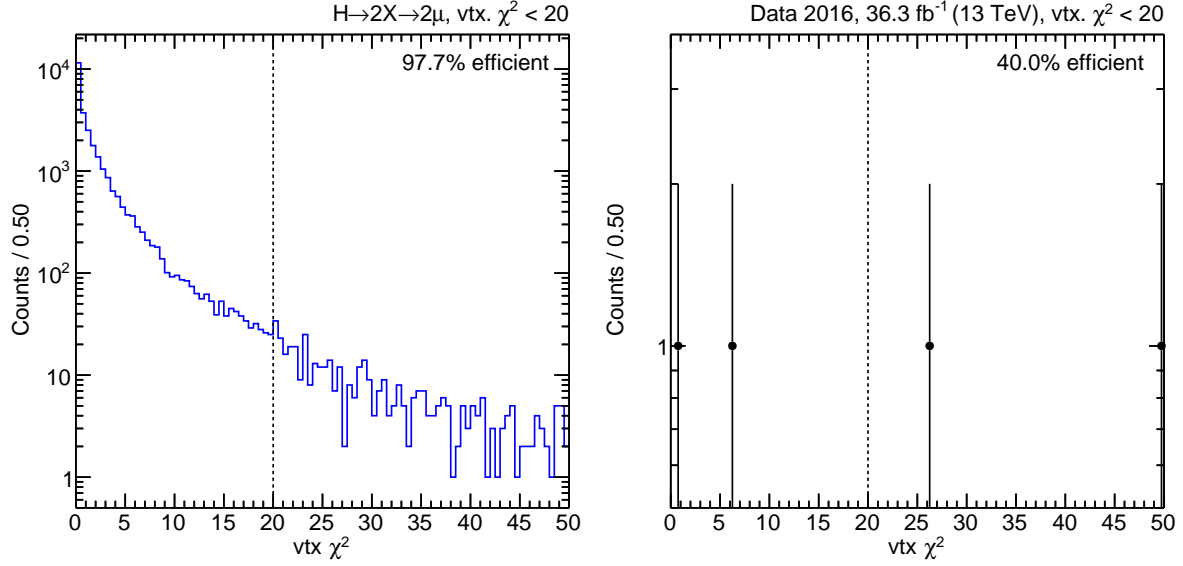


Figure 4.31: Histograms of events passing the full selection except for the vertex  $\chi^2$  cut, in (left)  $2\mu$  signal and (right) data in the control region, along with the labeled cut value (and corresponding dashed line) and the selection efficiency.

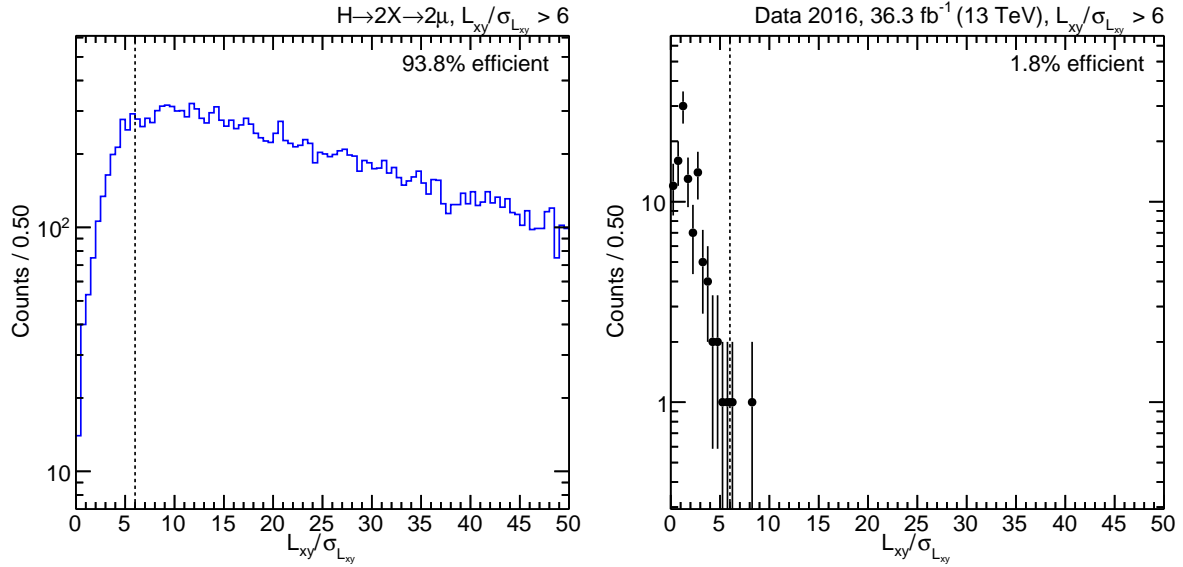


Figure 4.32: Histograms of events passing the full selection except for the  $L_{xy}/\sigma_{L_{xy}}$  cut, in (left)  $2\mu$  signal and (right) data in the control region, along with the labeled cut value (and corresponding dashed line) and the selection efficiency.

### 4.7.1 Luminosity Uncertainty

The uncertainty on the integrated luminosity for the 2016 data taking period is 2.5% [54]. This uncertainty directly applies to the signal estimate, as the number of signal events is estimated from Monte Carlo simulation and normalized to compare to the integrated luminosity in data. However, as the expected background is estimated from data, and not from Monte Carlo simulation, this uncertainty does not apply to the estimated background.

### 4.7.2 Background Uncertainties

The values of the transfer factors obtained in Section 4.5 were computed with somewhat arbitrary choices (e.g. for  $L_{xy}/\sigma_{L_{xy}}$  windows) and with assumptions that the values obtained from various control regions were relevant for estimating the background in the signal region. The associated uncertainties are expected to be the largest systematic uncertainties in the background estimation (though still smaller than uncertainties that are statistical in origin).

The value of the Drell-Yan background transfer factor varies within 2-7% as reasonable variations are made in the value of relative isolation satisfied by the PAT muons, in the  $L_{xy}/\sigma_{L_{xy}}$  values satisfied by the matched PAT dimuons, or across the mass windows containing the Z mass peak enumerated in Table 4.7. As the final results of the analysis are robust with respect to the assigned uncertainty, it is reasonable to assign a slightly larger number, a systematic uncertainty of 10%, to the Drell-Yan transfer factor  $\tau_{DY}$ .

The value of the QCD background transfer factor is dominated by statistical uncertainties arising from the small numbers of events used to estimate it. Changes to the value of the relative isolation satisfied by the PAT muons and to the  $L_{xy}/\sigma_{L_{xy}}$  values satisfied by the matched PAT dimuons have a small effect in comparison. Moreover, the value is found to depend on the mass windows, with larger values close to 10 GeV. The values of the QCD transfer factors are therefore computed in each mass window, with the systematic uncertainty defined to be the statistical uncertainty obtained by error propagation of the ratio of two Poisson counts used to estimate each factor. Table 4.8 enumerates the QCD transfer factors by mass window, along with their uncertainty.

Mass Window	${}^0_{\text{SS}}\text{CR}_{\text{QCD}}^{L>6}$	${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L>6}$	$\tau_{\text{QCD}}$	$\sigma_{\tau_{\text{QCD}}}$
$10 \text{ GeV} < m_{\mu\mu} < 32 \text{ GeV}$	103	302	$2.9 \pm 0.3$	11%
$20 \text{ GeV} < m_{\mu\mu} < 80 \text{ GeV}$	35	51	$1.5 \pm 0.3$	22%
$35 \text{ GeV} < m_{\mu\mu} < 245 \text{ GeV}$	10	18	$1.8 \pm 0.7$	39%
$65 \text{ GeV} < m_{\mu\mu}$	3	7	$2.3 \pm 1.6$	69%
Total	115	323	$2.8 \pm 0.9$	11%

Table 4.8: Transfer factors  $\tau_{\text{QCD}}$  for estimating QCD background, along with their systematic uncertainties, given in dimuon mass windows.

Events in data with  $L_{\text{xy}}/\sigma_{L_{\text{xy}}} < 6$  are used to cross check these transfer factors. The histograms of events in the control region  $\pi_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$  and in the control region  ${}^0_{\text{SS}}\text{CR}_{\text{full}}^{L<6}$  are scaled according to transfer factors obtained from the histograms of events in the control regions with  $L_{\text{xy}}/\sigma_{L_{\text{xy}}} > 6$  for estimating DY background and QCD background (respectively), and compared to the control region  ${}^0_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$ . These plots are given in Figure 4.33.

- The top left plot of Figure 4.33 shows the events passing the full selection in data for  ${}^0_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$  (labeled “SR”, colored red) and for  $\pi_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$  (labeled “CR”, colored blue), showing general agreement between the  $|\Delta\Phi| < \pi/4$  and  $|\Delta\Phi| > 3\pi/4$  regions.
- The binwise ratio of the  ${}^0_{\text{OS}}\text{CR}_{\text{DY}}^{L<6}$  and  $\pi_{\text{OS}}\text{CR}_{\text{DY}}^{L<6}$  histograms gives a weight factor for each  $L_{\text{xy}}/\sigma_{L_{\text{xy}}}$  bin; binwise scaling  $\pi_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$  by these weights gives the top right plot of Figure 4.33, with the modified histogram labeled “CR w/ DY” and colored green, to represent a correction by the Drell-Yan transfer factor.
- The binwise ratio of the  ${}^0_{\text{SS}}\text{CR}_{\text{QCD}}^{L<6}$  and  ${}^0_{\text{OS}}\text{CR}_{\text{QCD}}^{L<6}$  histograms gives a weight factor for each  $L_{\text{xy}}/\sigma_{L_{\text{xy}}}$  bin; binwise scaling the 4 events in  ${}^0_{\text{SS}}\text{CR}_{\text{full}}^{L<6}$  by these weights and adding them to the modified  $\pi_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$  histogram in the top right plot gives the bottom left plot of Figure 4.33, with the modified histogram labeled “CR w/ DY+QCD” and colored violet, to represent a correction by both the Drell-Yan transfer factor and the QCD transfer factor.

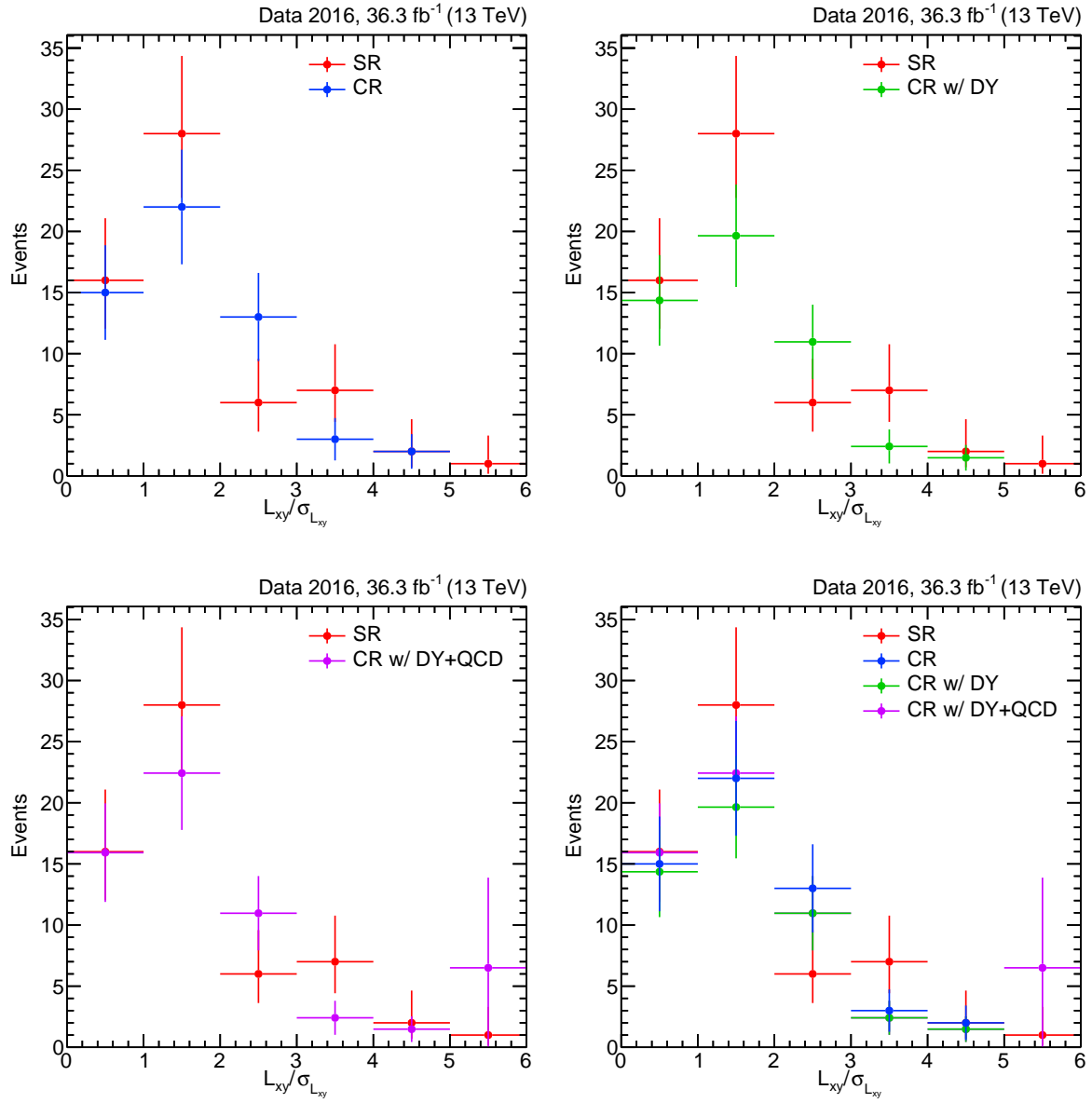


Figure 4.33: Histograms of  ${}^0_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$  and  ${}^{\pi}_{\text{OS}}\text{CR}_{\text{full}}^{L<6}$  in data, with the latter binwise scaled by Drell-Yan and QCD transfer factors, demonstrating agreement and validating the method of obtaining background estimates in the signal region with transfer factors from control regions. Error bars for the scaled histograms are computed approximately by linearized propagation of uncertainties of the quantities combined. Error bars for the SR histogram are frequentist (Garwood) 68% confidence level confidence intervals for the unknown Poisson mean.

All four histograms are plotted together in the bottom right plot of Figure 4.33 for comparison. The corrected “CR w/ DY+QCD” histogram (violet) has 59.7 events and binwise is consistent with the SR histogram, which has 60 events. This demonstrates consistency in the use of transfer factors from control regions to obtain estimates of background events in the signal region.

### 4.7.3 Signal Uncertainties

#### 4.7.3.1 Signal Lifetime Reweighting

As enumerated in Table 4.2, for a given BSM Higgs mass and long-lived particle mass  $m_H$  and  $m_X$ , this analysis simulated and produced signal samples for only three distinct values of  $c\tau$ . To calculate an estimated number of signal events for lifetimes other than these three nominal lifetimes, this analysis employs a procedure to reweight one signal lifetime into another. A particle with lifetime  $\tau$  will decay after a time  $t$  given by a decaying exponential probability distribution  $\mathcal{P}(t; \tau) = e^{-t/\tau}/\tau$ . A sample with a decay time distribution of reference lifetime  $\tau_{\text{ref}}$  can be reweighted so that it is distributed with a new lifetime  $\tau_{\text{new}}$  by multiplying the contribution of each event by a weight factor given by the ratio of the two distributions:

$$w_{\tau_{\text{ref}} \rightarrow \tau_{\text{new}}}(t) \equiv \frac{\mathcal{P}(t; \tau_{\text{new}})}{\mathcal{P}(t; \tau_{\text{ref}})} = \frac{\frac{1}{\tau_{\text{new}}} e^{-t/\tau_{\text{new}}}}{\frac{1}{\tau_{\text{ref}}} e^{-t/\tau_{\text{ref}}}} \quad (4.8)$$

where  $t$  can be computed from the generated values of suitable experimental quantities in a number of equivalent ways, including

$$t = \frac{L_{xy}}{m_X \cdot p_T^X} \quad (4.9)$$

When reweighting from longer (shorter) lifetimes to shorter (longer) lifetimes, events with short decay times are given large (small) weights and events with long decay times are given small (large) weights such that the reference decay time distribution (with unit area) is transformed into the new decay time distribution (also with unit area).

Let the number of generated signal events be  $N_{\text{gen}}$ , and let  $S$  be the subset of those  $N_{\text{gen}}$  events that pass the trigger and the full selection. With event weights, the estimated number



of passing events  $N_{\text{pass}}$  is

$$N_{\text{pass}} = \sum_{i \in S} w_{\tau_{\text{ref}} \rightarrow \tau_{\text{new}}}(t_i) \quad (4.10)$$

In the limit of large numbers of events, the sum of the event weights (for all generated events) for the reference and new lifetimes should be equal and equivalent to  $N_{\text{gen}}$ . In practice, instabilities associated with limited sample sizes and large weights generally result in

$$\sum_{i=0}^{N_{\text{gen}}} w_{\tau_{\text{ref}} \rightarrow \tau_{\text{new}}}(t_i) < N_{\text{gen}} \quad (4.11)$$

It is undesirable for the signal efficiency to suffer from the effect of this underestimated denominator, and so we are prompted to divide  $N_{\text{pass}}$  instead by the true number of generated events, rather than the sum of the weights. Thus, the signal selection efficiency is defined as

$$\varepsilon = \frac{N_{\text{pass}}}{N_{\text{gen}}} \quad (4.12)$$

Because the denominator of  $\varepsilon$  is an exactly known quantity, the binomial Clopper-Pearson interval does not apply to estimating the uncertainty on  $\varepsilon$ ; only the uncertainty on  $N_{\text{pass}}$  affects the uncertainty on  $\varepsilon$ . The uncertainty on the contribution of the  $i$ th event is its weight,  $w_i$ . Treating each uncertainty as independent and uncorrelated, the uncertainty on the sum of all the passing events is obtained from the linear approximation to the propagation of uncertainty:

$$\sigma_{N_{\text{pass}}} = \sqrt{\sum_{i \in S} w_{\tau_{\text{ref}} \rightarrow \tau_{\text{new}}}(t_i)^2} \quad (4.13)$$

This expression reduces to the Poisson uncertainty on the number of passing events  $\sqrt{N_{\text{pass}}}$  when the weights are all 1, i.e. without any lifetime reweighting. It therefore also accounts for the statistical uncertainty arising from estimating an efficiency from a finite sample of events. The relative uncertainty on  $\varepsilon$  can then be defined as the ratio

$$\frac{\sigma_{\varepsilon}}{\varepsilon} = \frac{\sigma_{N_{\text{pass}}}}{N_{\text{pass}}} \quad (4.14)$$

For the signal samples (nominally, without lifetime reweighting) for which the analysis selection gives non-negligible efficiency (which are the medium and long lifetime samples), the relative uncertainty (which is due to statistical uncertainty only) is 2–5%. The uncertainty

due to lifetime reweighting increases when reweighting the short and medium lifetime samples to much smaller lifetimes (for which the analysis selection gives negligible efficiency) and when reweighting the short and medium lifetime samples to longer lifetimes (because there are not enough events at intermediate lifetimes to sufficiently estimate the efficiency, and what events there are have large weights). For the purposes of calculating upper limits, the relative uncertainty (Equation 4.14) is computed for each (reweighted and nominal) signal sample and taken as a systematic uncertainty on the estimated number of signal events. Any signal estimates with relative uncertainty greater than 50% are omitted from the calculations.

#### 4.7.3.2 Pileup Reweighting

The mean number of pp collisions per bunch crossing is a measure of the phenomenon known as pileup. Events with more pileup produce more particles and more tracks in the tracker and in the muon system, which directly impact e.g. the reconstruction efficiencies, or the efficiency to associate DSA muons with PAT muons. The distribution of pileup is different in data than in simulation, and so the distribution in signal simulation is reweighted to correspond to the observed distribution in data. This is performed by producing a binned histogram of the distribution of the number of true (as opposed to reconstructed) primary vertices in each event (an estimate of the pileup) for 2016 data and for signal simulation, scaling the histograms to unit area, and dividing the frequency in each bin of the data distribution by the frequency in each bin of the signal distribution.

Figure 4.34 shows the distribution of pileup in data and in simulation. The black dots are the (normalized) distribution of pileup in CMS data taken in 2016 certified for muon physics. The red dots are the (normalized) distribution of pileup in all  $2\mu$  signal samples combined. To confirm that this combined distribution is robust with respect to signal parameters, the orange bands represent the binwise maximum and binwise minimum of all 33 sets of signal parameters. The right plot of Figure 4.34 shows the ratio of the data distribution to the signal distribution(s) in the left plot. Each bin—corresponding to a number of true primary vertices—now contains the signal event weight.

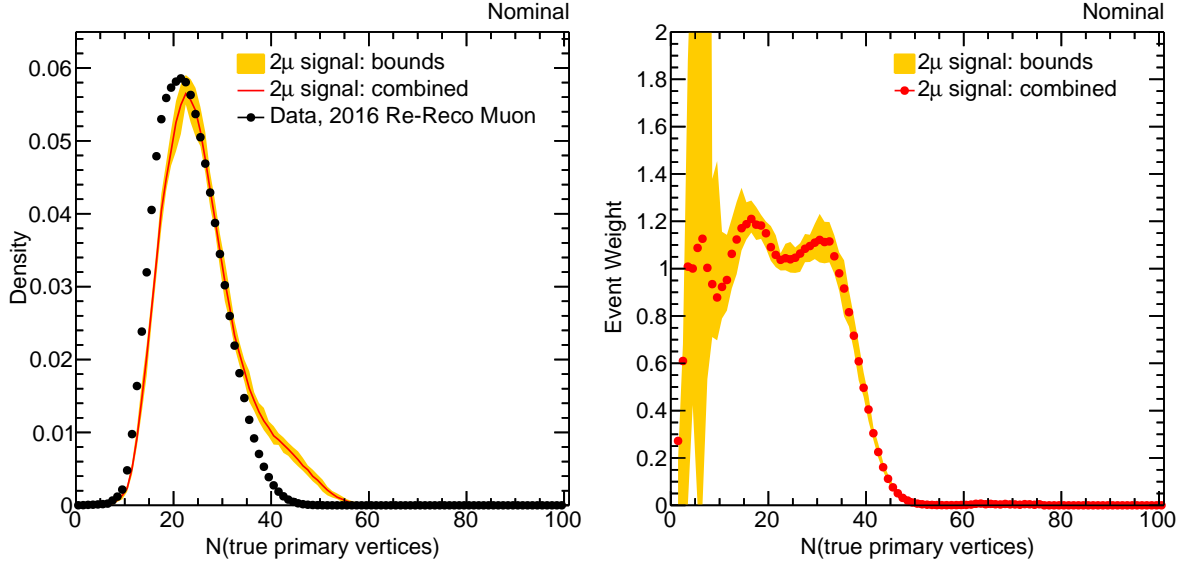


Figure 4.34: (Left) Histogram of the number of true primary vertices in 2016 data and in  $2\mu$  signal. (Right) Ratio of the data histogram to the signal histogram, giving the nominal signal event weight for a given number of true primary vertices.

To formalize the notation, let  $\mathcal{D}(P)$  be the histogram in data (normalized to unit area), and  $\mathcal{S}(P)$  be the the histogram for all signal samples combined (normalized to unit area), giving the distribution of pileup  $P$ . Let  $P(i)$  be the number of true primary vertices in the  $i$ th simulated signal event. Then the pileup weight factor for the  $i$ th signal event is

$$w_{\text{pileup}}(P(i)) = \frac{\mathcal{D}(P)}{\mathcal{S}(P)} \quad (4.15)$$

and the contribution of each event is scaled by the weight factor.

There is a systematic uncertainty associated with this pileup reweighting procedure arising from an uncertainty in the minimum-bias pp cross section used to calculate the data pileup distribution; the nominal value of this cross section is 69.2 mb. To estimate the effect of this uncertainty, the pp cross section used to calculate the pileup distribution in data is varied by  $\pm 5\%$ , and the analysis signal yield is recomputed using the pileup event weights obtained from the upwards variation and downwards variation. For signal samples for which the analysis selection gives non-negligible efficiency, the variation due to variations in pileup

Uncertainty	Value	Signal	Background
Lifetime reweighting	2–50%	✓	
Luminosity	2.5%	✓	
Pileup	2%	✓	
Trigger efficiency	15%	✓	
Reconstruction efficiency	15%	✓	
Transfer factor $\tau_{\text{DY}}$	10%		✓
Transfer factor $\tau_{\text{QCD}}$	11–69%		✓
Statistical $\pi_{\text{OS}} \text{CR}_{\text{full}}^{L>6} \rightarrow \text{SR}$	see Section 4.8.1		✓

Table 4.9: Summary of systematic uncertainties in this analysis, along with a check mark indicating whether the uncertainty applies to the signal or to the background estimate. As discussed in Section 4.8.1, the statistical uncertainty due to extrapolating a background estimate in a signal region from a control region is treated specially, and represents the dominant systematic uncertainty for the background estimate.

reweighting was found to be 2% or less for all signal samples. Therefore, the systematic uncertainty on the signal estimate due to pileup modeling is taken to be 2%.

The definitions of  $N_{\text{pass}}$ ,  $N_{\text{gen}}$ , and  $\sigma_{N_{\text{pass}}}$  in Equations 4.10–4.14 are consequently modified to be

$$N_{\text{pass}} \rightarrow \sum_{i \in S} [w_{\tau_{\text{ref}} \rightarrow \tau_{\text{new}}}(t_i) \cdot w_{\text{pileup}}(P(i))] \quad (4.16)$$

$$N_{\text{gen}} \rightarrow \sum_{i=0}^{N_{\text{gen}}} w_{\text{pileup}}(P(i)) \quad (4.17)$$

$$\sigma_{N_{\text{pass}}} \rightarrow \sqrt{\sum_{i \in S} [w_{\tau_{\text{ref}} \rightarrow \tau_{\text{new}}}(t_i) \cdot w_{\text{pileup}}(P(i))]^2} \quad (4.18)$$

As the pileup weights are small compared to the weights due to signal lifetime reweighting, the uncertainty on the modified  $N_{\text{gen}}$  is taken to be negligible, and the procedure for obtaining the uncertainty on the signal estimate due to lifetime reweighting as described in Section 4.7.3.1 is not modified, except for the definitions in Equations 4.16–4.18.

### 4.7.3.3 Other Signal Systematic Uncertainties

The efficiencies of the trigger and of the DSA reconstruction are different in data than in simulation. Estimating them in data is challenging because a real source of displaced muons is required; a dedicated study for estimating these efficiencies from data using cosmic ray muons with large impact parameters is in progress. Other than potentially large uncertainties from lifetime reweighting, these systematic uncertainties are expected to be the dominant systematic uncertainties on the estimated signal. Based on our current understanding, each of the trigger and reconstruction efficiencies are assigned a systematic uncertainty of 15%. Other systematic uncertainties, such as theoretical uncertainties from next-to-leading-order modeling and choices of parton distribution functions, are taken to be negligible for this analysis.

## 4.8 Results

### 4.8.1 Computing Upper Limits with the $\text{CL}_s$ Criterion

The procedure for setting an upper limit on the long-lived particle production cross section times the branching fraction to two muons is as follows [55]. Let the probability of obtaining the observed data  $x$  be  $\mathcal{P}(x; \mu, \theta)$  (probability density function, or pdf, if  $x$  is continuous). Here  $\mu$  is the signal strength, a quantity proportional to the signal cross section, so that  $\mu = 0$  represents a background-only model, and  $\theta$  represents the systematic uncertainties, also referred to as nuisance parameters. A model therefore defines an expected number of signal events, an expected number of background events, and a set of systematic uncertainties. Then the likelihood function for a model with signal strength  $\mu$  and nuisance parameters  $\theta$  given some observed data  $x = X$  is  $\mathcal{L}(\mu, \theta | X)$ . This likelihood function is globally maximized at some values  $\mu = \hat{\mu}$  and  $\theta = \hat{\theta}$ , and for a specific value of  $\mu$ , locally maximized at  $\theta = \hat{\theta}_\mu$ . The LHC-style test statistic [56] is then defined, based on a ratio of profile likelihoods, as

$$q_{\text{LHC}}(\mu) = -2 \ln \left[ \frac{\mathcal{L}(\mu, \hat{\theta}_\mu | X)}{\mathcal{L}(\mu = \hat{\mu}, \theta = \hat{\theta} | X)} \right] \quad (4.19)$$

The likelihoods are evaluated by treating the signal strength in a frequentist manner and incorporating the systematic uncertainties in a fully frequentist manner by profiling them. Limits were also calculated by marginalizing (or integrating over) the systematic uncertainties, following the Cousins-Highland prescription of treating the uncertainties in a Bayesian manner [57]. The upper limits were in general a little lower, which can be attributed to marginalization of the mean number of background events, leading to some mild under-coverage when the true background mean is small. Thus the profiling of nuisance parameters is preferred.

For most types of systematic uncertainties, the pdfs are assumed to follow a log-normal distribution, with one exception. There is a systematic uncertainty on the estimated background associated with the statistical uncertainty arising from estimating the expected background in the signal region from a number of observed events in the control region. The pdf for this systematic uncertainty is assumed to follow a gamma distribution, which has good frequentist properties; from a Bayesian point of view, it is the posterior pdf for the background mean if the prior pdf is uniform and the likelihood function is from one sample of a Poisson distribution [53, 58].

For some observation  $X$ , the minimum value  $Q(X)$  of the test statistic  $q_{\text{LHC}}(\mu)$  corresponds to the value of  $\mu$  with maximum likelihood. Then the pdf of the minimum value of the test statistic  $\mathcal{P}(Q; \mu)$ , given a signal strength  $\mu$ , is computed, by sampling observations  $X$  from ensembles of toy experiments generated using Monte Carlo methods, computing the likelihood functions, and minimizing the test statistic. The actual observation  $X_{\text{obs}}$  in CMS data taken in 2016 corresponds to some minimum value of the test statistic  $Q_{\text{obs}} = Q(X_{\text{obs}})$ . Define the  $p$ -values for signal and background as

$$p_\mu = \int_{-\infty}^{Q_{\text{obs}}} \mathcal{P}(Q; \mu) dQ \quad (4.20)$$

$$1 - p_b = \int_{-\infty}^{Q_{\text{obs}}} \mathcal{P}(Q; 0) dQ \quad (4.21)$$

In the 1990s and decades before, using the standard  $p$ -value  $p_\mu$  to place upper limits on  $\mu$  was known to cause difficulties if there was a severe downward fluctuation in the background. Three proposed solutions became widely accepted and are described in the PDG Review of Particle Physics statistical section [59], namely, the method advocated by Feldman and Cousins; a Bayesian method with priors that are known to give good frequentist coverage or over-coverage; and an invention in high-energy physics known as  $\text{CL}_s$ . By convention the ATLAS and CMS collaborations use  $\text{CL}_s$  [60, 61] for most results, and this convention is followed here. It leads to upper limits that over-cover at the stated confidence level, if the statistical model is correct. This modified frequentist statistic  $\text{CL}_s$  is defined as

$$\text{CL}_s = \frac{p_\mu}{1 - p_b} \quad (4.22)$$

For some confidence level  $1 - \alpha$ , signal models with values of  $\text{CL}_s$  such that  $\text{CL}_s \leq \alpha$  are excluded at a confidence level of  $1 - \alpha$ . The 95% confidence level upper limit on the signal strength  $\mu_{\text{UL}}$  is therefore the value of  $\mu$  such that  $\text{CL}_s = 0.05$ .

#### 4.8.2 Technical Details of Using the Higgs Combine Tool

The CMS software framework includes a statistical analysis package known as the Higgs Combine tool (or simply `combine`) used throughout CMS for computing statistical significance and upper limits and for performing goodness-of-fit tests [56]. The choices of statistical methods described above were made because this analysis involves low numbers of events in both the signal and control regions. These choices result in a technical procedure for using `combine` to compute upper limits that differs from that used by many CMS analyses. For this reason, the procedure and the command-line options to `combine` are documented here.

Using `combine` begins with providing an expected number of signal events (given a model), an expected number of background events, (nominal) values for the systematic uncertainties along with the functional form of their distributions, and an observed number of events in data. The `combine` program is then run with the following configurations controlling the statistical treatment:

- `--rule CLs --cl 0.95`: computes the  $\text{CL}_s$  upper limit at a 95% confidence level.
- `--method HybridNew`: generates toy experiments to compute the distribution of the test statistic, instead of using the asymptotic frequentist approximations of the distribution valid only for large numbers of events.
- `--toysH 5000`: generates 5000 toy experiments each iteration.
- `--iterations 1`: generates `toysH` toy experiments 1 time.
- `--seed -1`: randomizes the seed
- `--hintMethod AsymptoticLimits`: estimate the upper limit on the signal strength using asymptotic frequentist methods, and use the result to inform the range of signal strengths over which to search when running toy experiments.
- `--testStat LHC`: compute the distribution of the LHC-style test statistic.
- `--generateNuisances 0 --generateExternalMeasurements 1 --fitNuisances 1`: treats the nuisance parameters in a frequentist manner by profiling them. For each toy experiment, the value of the nuisance parameters are fixed to their post-fit values, making the method sensitive to observations in data.
- `--expectedFromGrid [QUANTILE]`: computes the QUANTILE quantile of the expected upper limit. When omitted, the program computes the observed limit; otherwise, QUANTILE is 0.5 for the median of the distribution under the background-only hypothesis; 0.16 and 0.84 for the lower and upper edges of the 68% quantile; and 0.025 and 0.975 for the lower and upper edges of the 95% quantile.

To provide `combine` with an expected number of  $2\mu$  signal events, given a signal model defined by values for  $m_H$ ,  $m_X$ , and  $c\tau$ , let

$$N_s^{\text{exp}} = [\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)]_{\text{ass}} \cdot \mathcal{L}_{\text{int}} \cdot \varepsilon \quad (4.23)$$

where  $\varepsilon$  was defined in Sections 4.7.3.1–4.7.3.2, and  $[\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)]_{\text{ass}}$  is an assumed arbitrary value of the  $H \rightarrow XX$  production cross section times the branching ratio for the



decay of the long-lived particle to two muons  $X \rightarrow \mu\mu$ . Then `combine` returns  $\mu_{\text{UL}}$ , which can be interpreted as the ratio of the upper limit on the number of  $2\mu$  signal events to the expected number of  $2\mu$  signal events:

$$\mu_{\text{UL}} = \frac{N_s^{\text{UL}}}{N_s^{\text{exp}}} = \frac{[\sigma(\text{H} \rightarrow \text{XX})B(X \rightarrow \mu\mu)]_{\text{UL}} \cdot \mathcal{L}_{\text{int}} \cdot \varepsilon}{[\sigma(\text{H} \rightarrow \text{XX})B(X \rightarrow \mu\mu)]_{\text{ass}} \cdot \mathcal{L}_{\text{int}} \cdot \varepsilon} = \frac{[\sigma(\text{H} \rightarrow \text{XX})B(X \rightarrow \mu\mu)]_{\text{UL}}}{[\sigma(\text{H} \rightarrow \text{XX})B(X \rightarrow \mu\mu)]_{\text{ass}}} \quad (4.24)$$

Therefore

$$[\sigma(\text{H} \rightarrow \text{XX})B(X \rightarrow \mu\mu)]_{\text{UL}} = \mu_{\text{UL}} \cdot [\sigma(\text{H} \rightarrow \text{XX})B(X \rightarrow \mu\mu)]_{\text{ass}} \quad (4.25)$$

### 4.8.3 Events Passing the Full Selection Criteria

8 opposite-sign events and 4 same-sign events pass the full selection in 2016 data. Table 4.10 enumerates these events, with their identifying run, lumi section, and event numbers, and their values of  $m_{\mu\mu}$ ,  $|\Delta\Phi|$ , and  $L_{xy}/\sigma_{L_{xy}}$ . The table also contains references to the event displays for the opposite-sign events in Figures 4.35–4.42. As in Figure 4.15, views are presented in both the transverse plane ( $\rho$ - $\phi$ ) and the orthogonal  $\rho$ - $z$  plane. DSA muon tracks are drawn in yellow (associated segments drawn as yellow dots); PAT muons are drawn in red; jets are drawn in cyan; tracks in the tracker are drawn in gray; missing transverse energy is drawn in purple. Detailed inspection of the individual events passing all cuts gives no indication that the events are not individually compatible with known backgrounds. As the numbers of events passing all cuts are compatible with expectations based on the background estimates and the transfer factors, we conclude that no excess is observed.

### 4.8.4 Upper Limits on the Cross Section Times Branching Ratio

Table 4.11 enumerates the number of opposite-sign events in the control region  $\pi_{\text{OS}}\text{CR}_{\text{full}}^{L>6}$ , the number of same-sign events in the control region  $^0_{\text{SS}}\text{CR}_{\text{full}}^{L>6}$ , the corresponding number of expected background events in the signal region, and the number observed events in the signal region, for each mass window in Table 4.7, given the events enumerated in Table 4.10. Given these observations, the systematic uncertainties described in Section 4.7, and the procedure described in Sections 4.8.1–4.8.2, upper limits on the signal strength for each mass

hypothesis are computed and converted to limits on the cross section times the branching ratio. Figures 4.43–4.46 are plots of the upper limits on  $\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)$  for various values of BSM Higgs mass  $m_H$  and long-lived particle mass  $m_X$  as a function of  $c$  times the long-lived particle mean proper lifetime,  $c\tau$ . In the figure for each limit curve, there is also shown the median upper limit that would be obtained in an imagined ensemble of similar experiments having only background events. The spread of upper limits about the median in such an ensemble of background-only experiments is indicated by yellow bands containing the central 68% quantile. Table 4.12 summarizes the values of the median and observed upper limits for the 33 sets of signal parameters ( $m_H$ ,  $m_X$ ,  $c\tau$ ) that do not require lifetime reweighting. Some values are omitted, corresponding to signal samples with zero efficiency or statistical uncertainty greater than 50%.

#### 4.8.5 Robustness of Upper Limits to Systematic Uncertainties

The computation and values of the upper limits on the signal strength are expected to be robust with respect to the values of the systematic uncertainties, dominated instead by uncertainties that are statistical in origin. This is verified with a study using a single point and testing the effect of varying the inputs in three different ways.

- **Test #1.** Fixing the signal systematic uncertainty at 20% and varying the uncertainty on the background transfer factor uncertainties from 5% to 40% in steps of 5% worsens the upper limit (increases its value) by 5%.
- **Test #2.** Fixing the background transfer factor uncertainty at 20% and changing the signal systematic uncertainty from 20% to 5% improves the upper limit (decreases its value) by 5%.
- **Test #3.** Changing the number of observed and expected background events from 3 to 0 improves the upper limit (decreases its value) by a factor of 2.

Large effects on the upper limits are therefore caused by changes to the background estimates, and not by changes to the precise values of the systematic uncertainties on those estimates.

Opposite-Sign Events in $ \Delta\Phi  < \pi/4$						
Run	Lumi	Event	$m_{\mu\mu}$ [GeV]	$ \Delta\Phi $	$L_{xy}/\sigma_{L_{xy}}$	Display
274969	944	1675672448	10.2	0.0764	56.91	Figure 4.35
276935	522	801995838	24.0	0.1309	6.76	Figure 4.36
276940	97	59821031	14.8	0.1123	14.22	Figure 4.37
276947	4	6168006	15.4	0.0020	6.73	Figure 4.38
277168	114	203474554	24.6	0.0714	76.68	Figure 4.39
278017	169	188965749	64.4	0.3114	8.90	Figure 4.40
278308	390	635154343	184.1	0.3784	8.81	Figure 4.41
281797	501	676135274	15.6	0.0468	87.70	Figure 4.42

Same-Sign Events in $ \Delta\Phi  < \pi/4$						
Run	Lumi	Event	$m_{\mu\mu}$ [GeV]	$ \Delta\Phi $	$L_{xy}/\sigma_{L_{xy}}$	
276935	641	1009966664	32.0	0.0327	57.6	
277070	717	1337738509	13.5	0.0555	6.2	
278240	240	484842145	10.6	0.1733	23.6	
279767	774	1271651549	13.4	0.4288	106.0	

Table 4.10: List of events passing the full selection in 2016 data, with their run, lumi section, and event numbers, the values of  $m_{\mu\mu}$ ,  $|\Delta\Phi|$ , and  $L_{xy}/\sigma_{L_{xy}}$ , and the associated event display.

Dimuon Mass Window	O.S. in $\pi_{\text{OS}}\text{CR}_{\text{full}}^{L>6}$	S.S. in ${}^0_{\text{SS}}\text{CR}_{\text{full}}^{L>6}$	Expected	Observed
$10 \text{ GeV} < m_{\mu\mu} < 32 \text{ GeV}$	0	3	8.8	6
$20 \text{ GeV} < m_{\mu\mu} < 80 \text{ GeV}$	0	1	1.5	3
$35 \text{ GeV} < m_{\mu\mu} < 245 \text{ GeV}$	0	0	0.0	2
$65 \text{ GeV} < m_{\mu\mu}$	0	0	0.0	1
Total	0	4	11.2	8

Table 4.11: Number of opposite-sign events in  $\pi_{\text{OS}}\text{CR}_{\text{full}}^{L>6}$ , same-sign events in  ${}^0_{\text{SS}}\text{CR}_{\text{full}}^{L>6}$ , expected background events in SR, and observed events in SR.

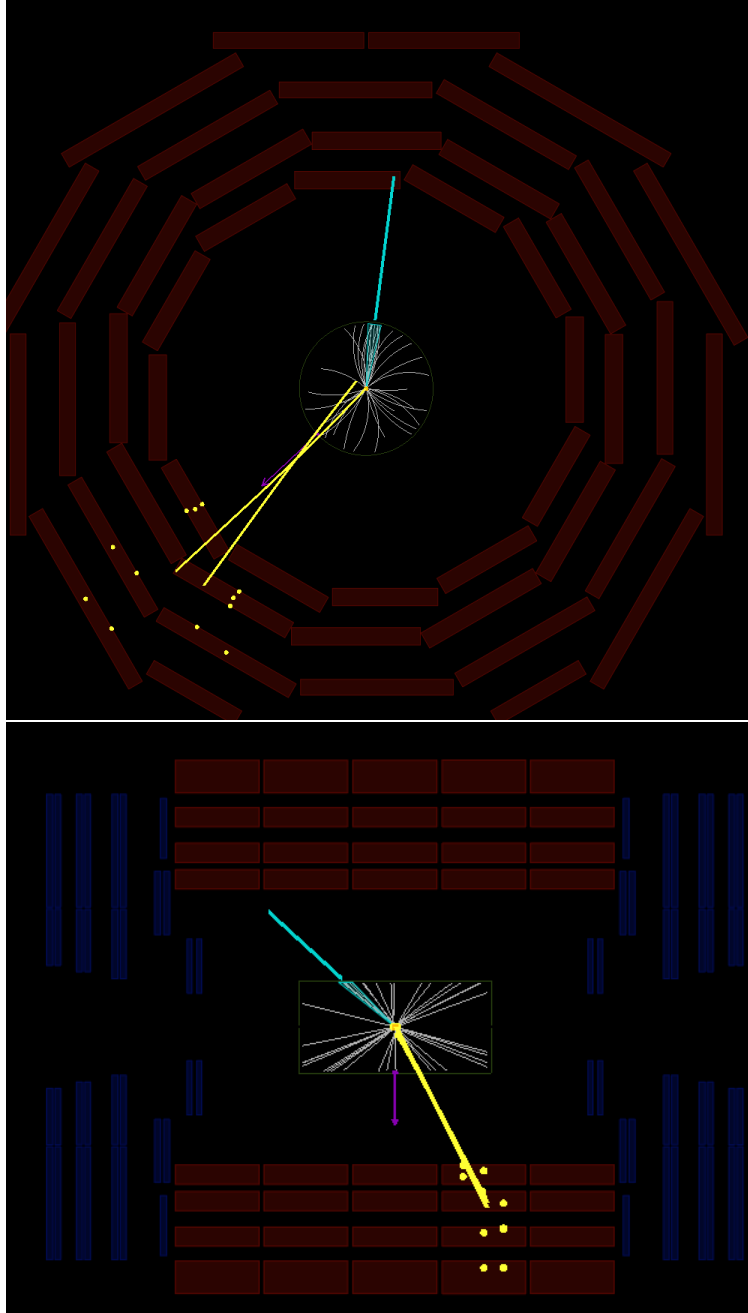


Figure 4.35: Display of  $\rho$ - $\phi$  and  $\rho$ - $z$  views of event 1675672448 in 2016 data, in run 274969, lumi section 944.

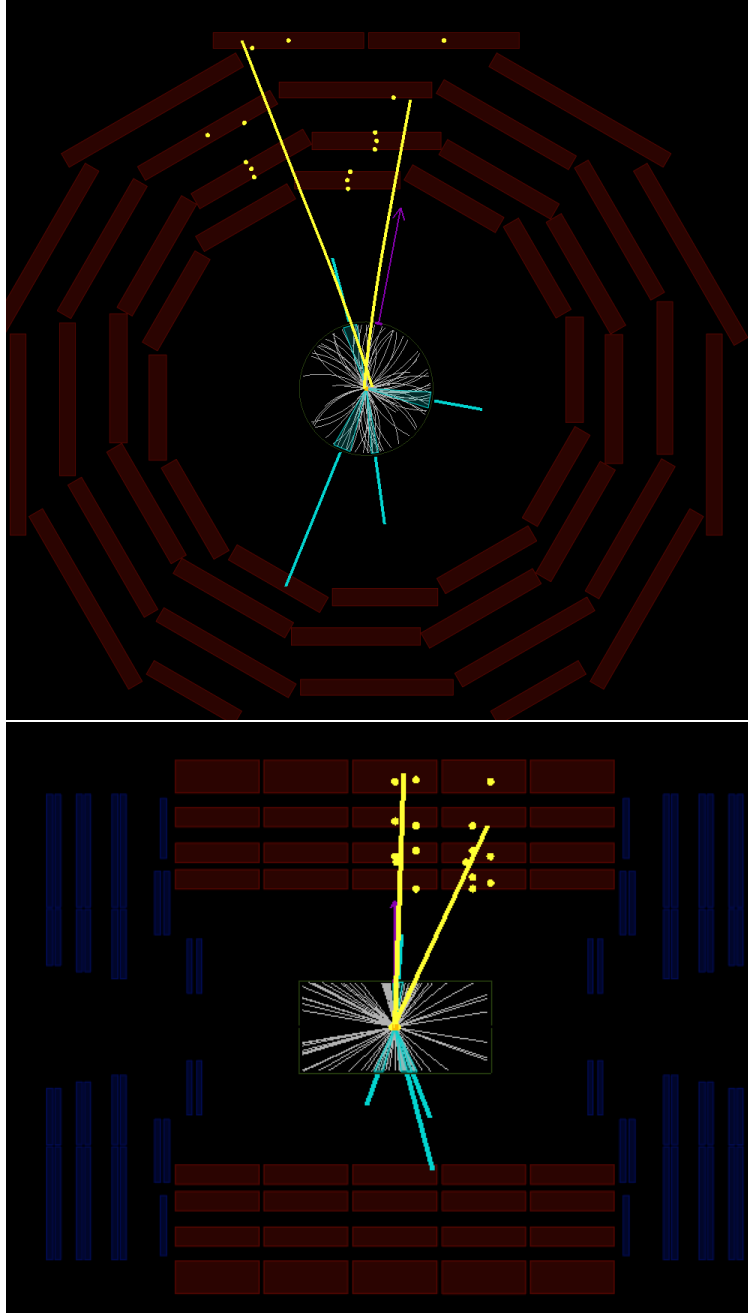


Figure 4.36: Display of  $\rho$ - $\phi$  and  $\rho$ - $z$  views of event 801995838 in 2016 data, in run 276935, lumi section 522.

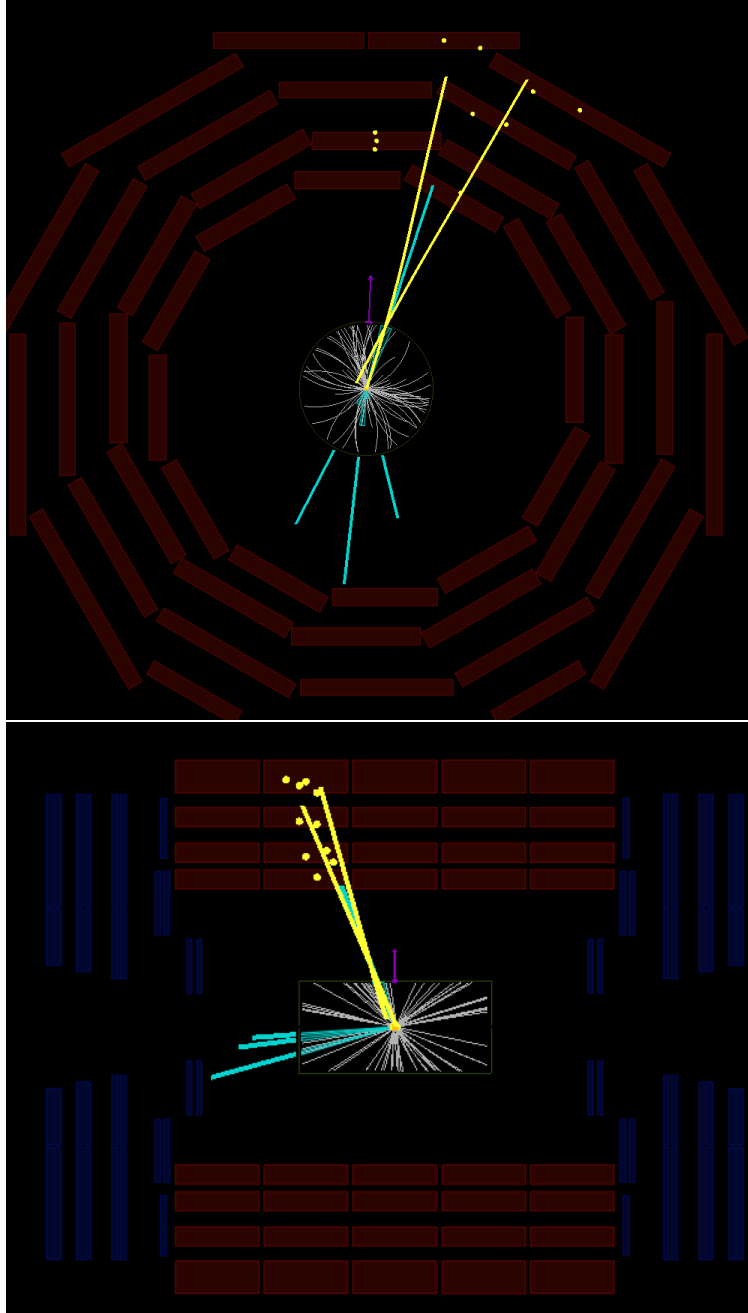


Figure 4.37: Display of  $\rho$ - $\phi$  and  $\rho$ - $z$  views of event 59821031 in 2016 data, in run 276940, lumi section 97.

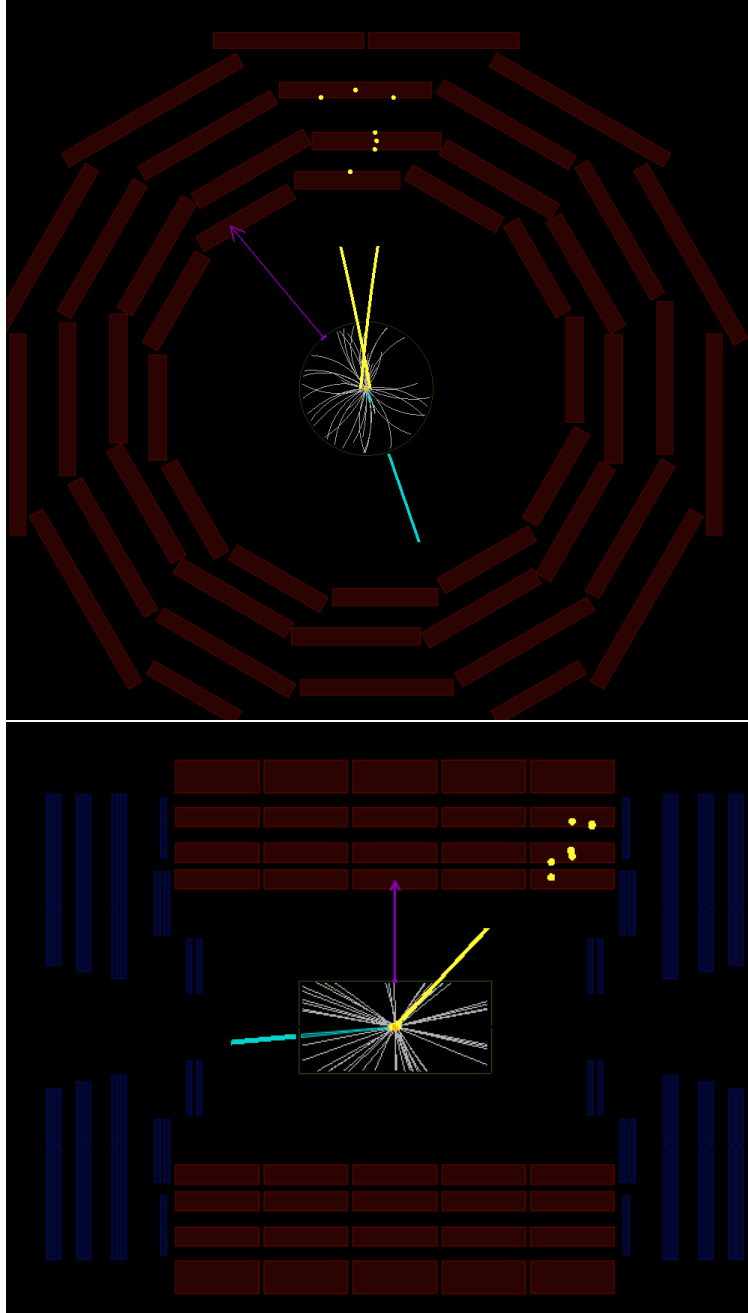


Figure 4.38: Display of  $\rho$ - $\phi$  and  $\rho$ - $z$  views of event 6168006 in 2016 data, in run 276947, lumi section 4.

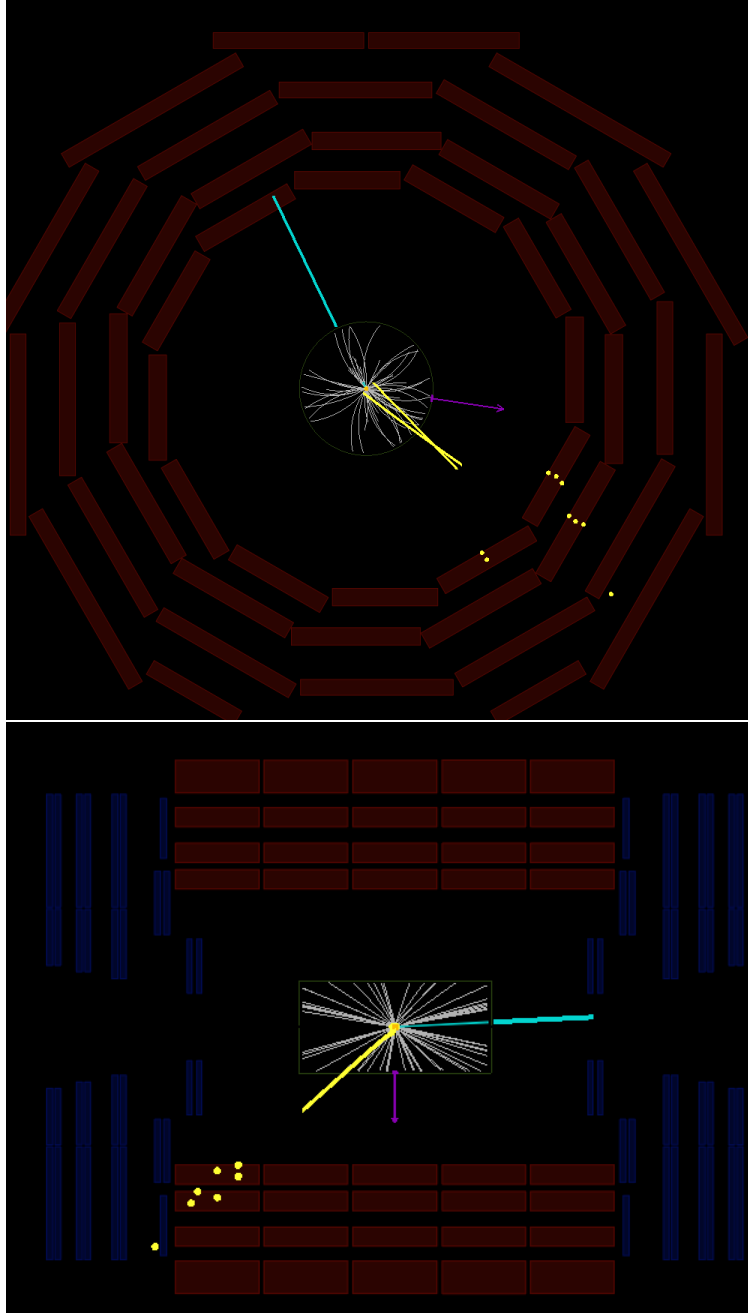


Figure 4.39: Display of  $\rho$ - $\phi$  and  $\rho$ - $z$  views of event 203474554 in 2016 data, in run 277168, lumi section 114.



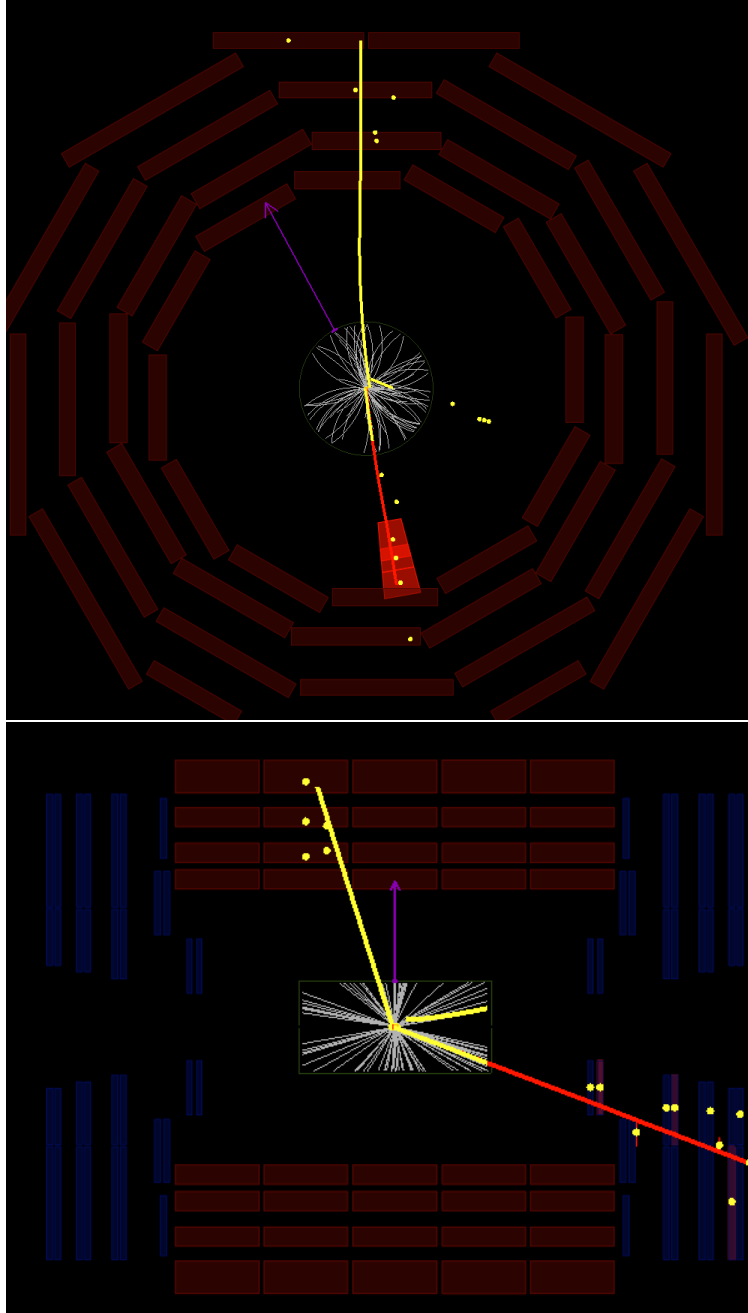


Figure 4.40: Display of  $\rho$ - $\phi$  and  $\rho$ - $z$  views of event 188965749 in 2016 data, in run 278017, lumi section 169.

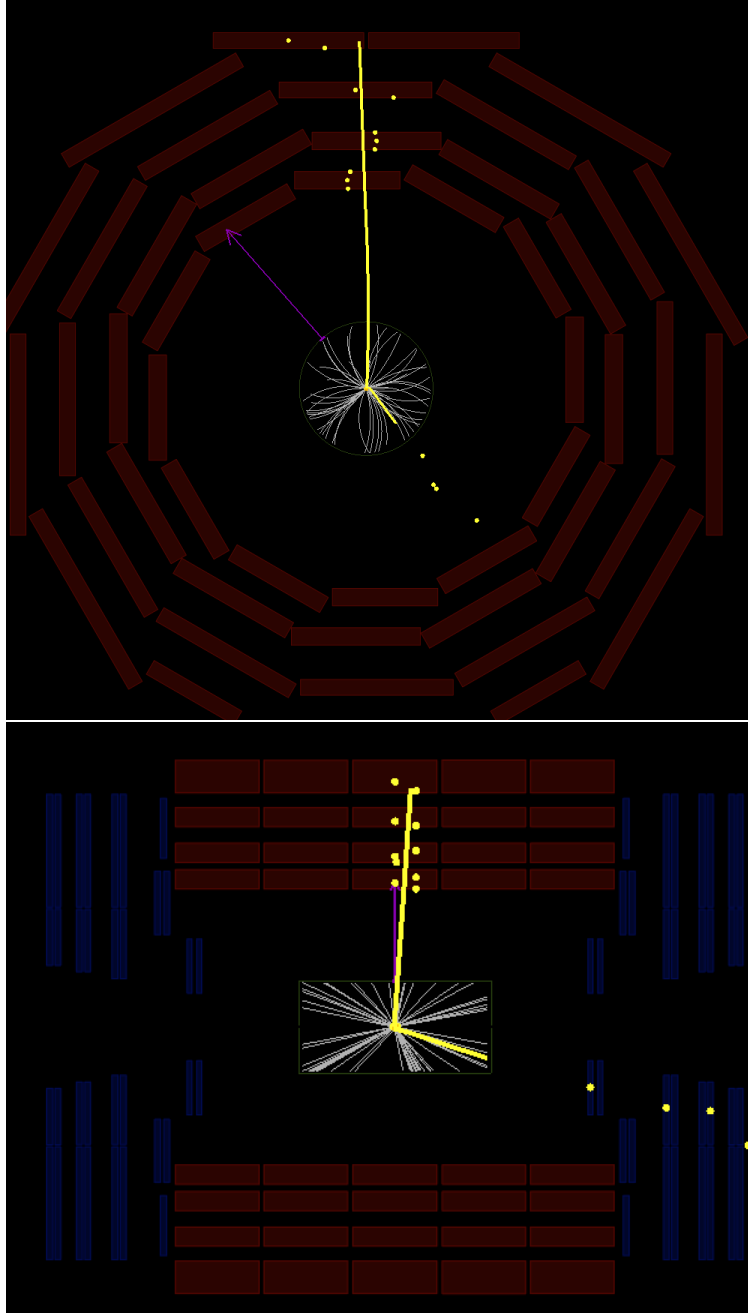


Figure 4.41: Display of  $\rho$ - $\phi$  and  $\rho$ - $z$  views of event 635154343 in 2016 data, in run 278308, lumi section 390.

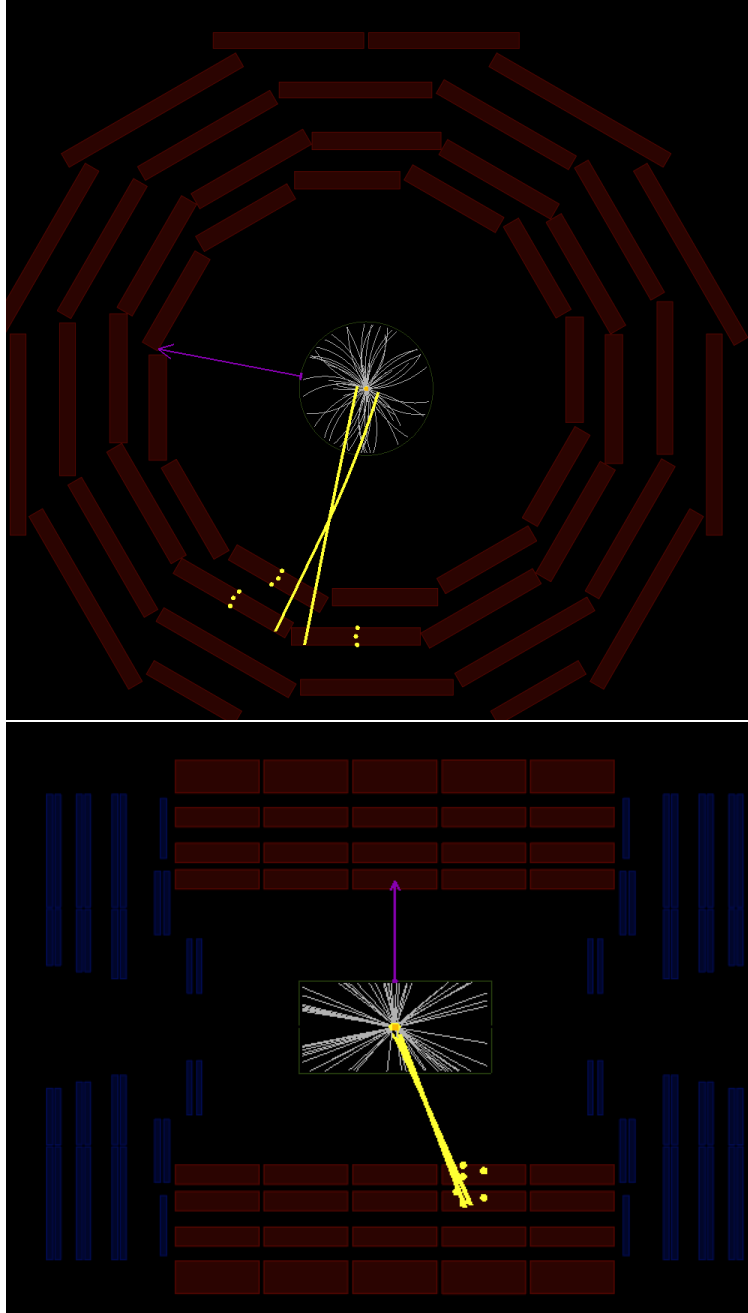


Figure 4.42: Display of  $\rho$ - $\phi$  and  $\rho$ - $z$  views of event 676135274 in 2016 data, in run 281797, lumi section 501.

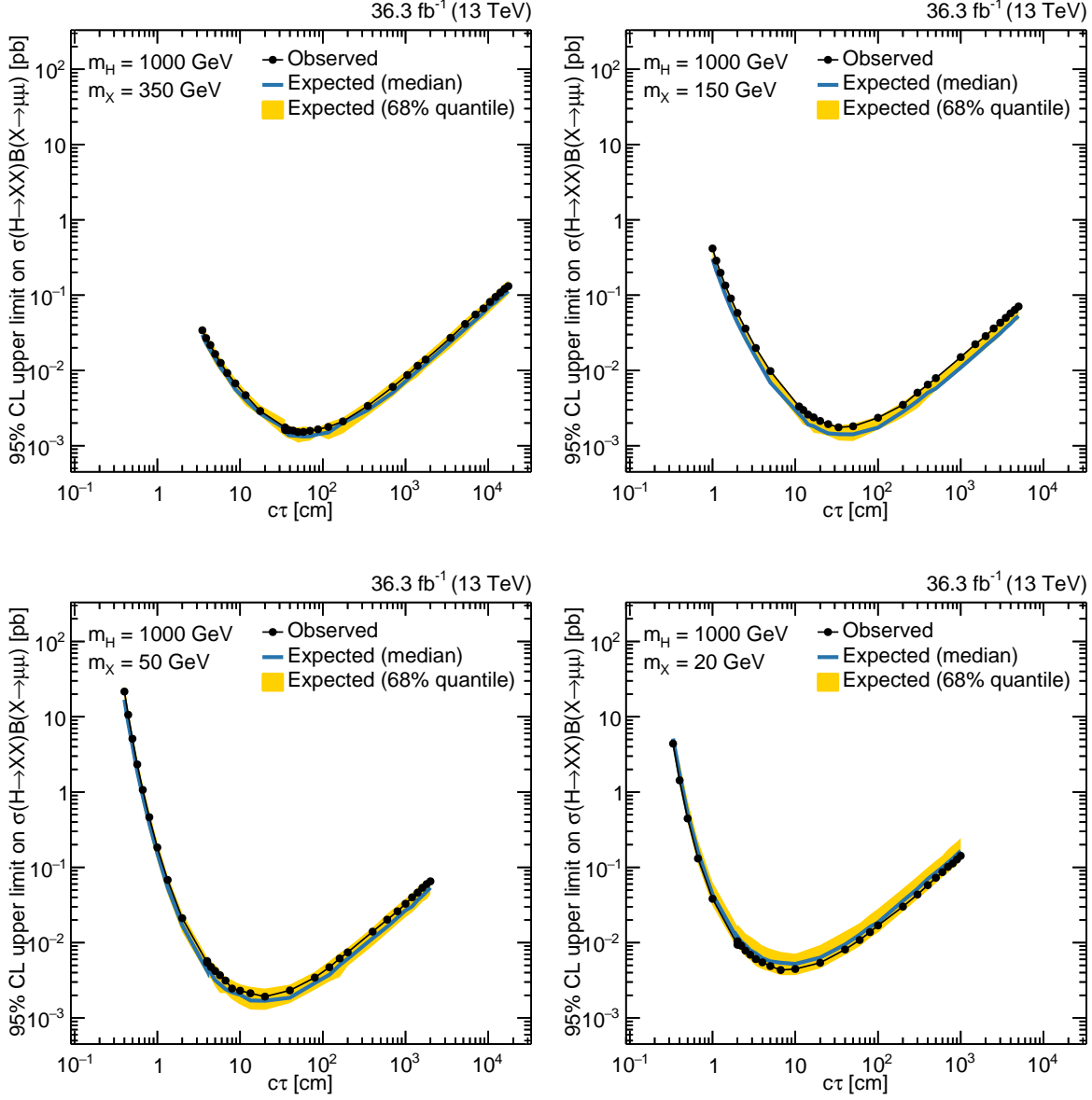


Figure 4.43: 95% CL upper limits on  $\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)$  for  $m_H = 1000$  GeV for various values of  $m_X$ . Black dots are observed limits; the blue line represents the median expected limits; the yellow shaded bands show the central 68% quantile.

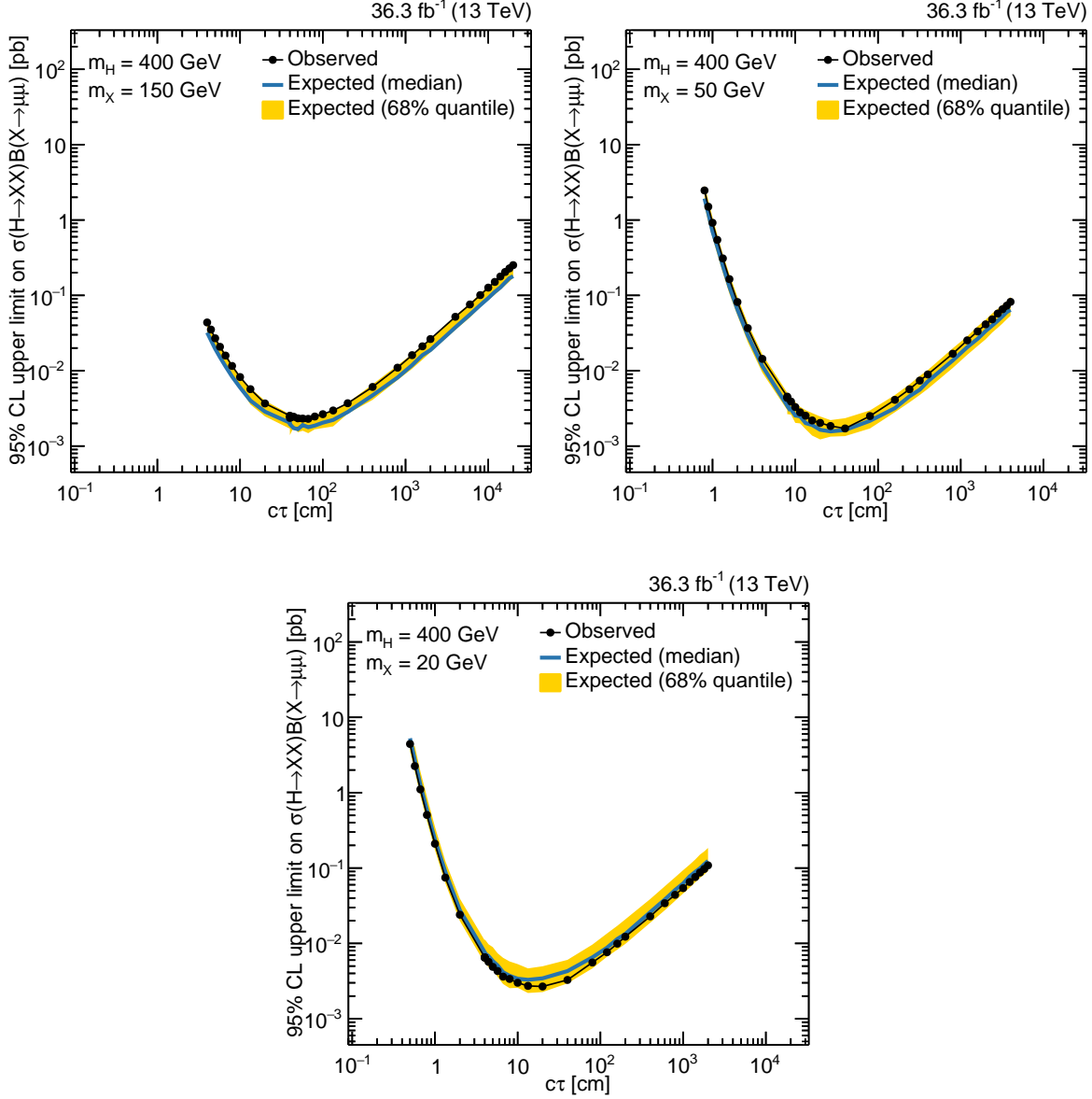


Figure 4.44: 95% CL upper limits on  $\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)$  for  $m_H = 400$  GeV for various values of  $m_X$ . Black dots are observed limits; the blue line represents the median expected limits; the yellow shaded bands show the central 68% quantile.

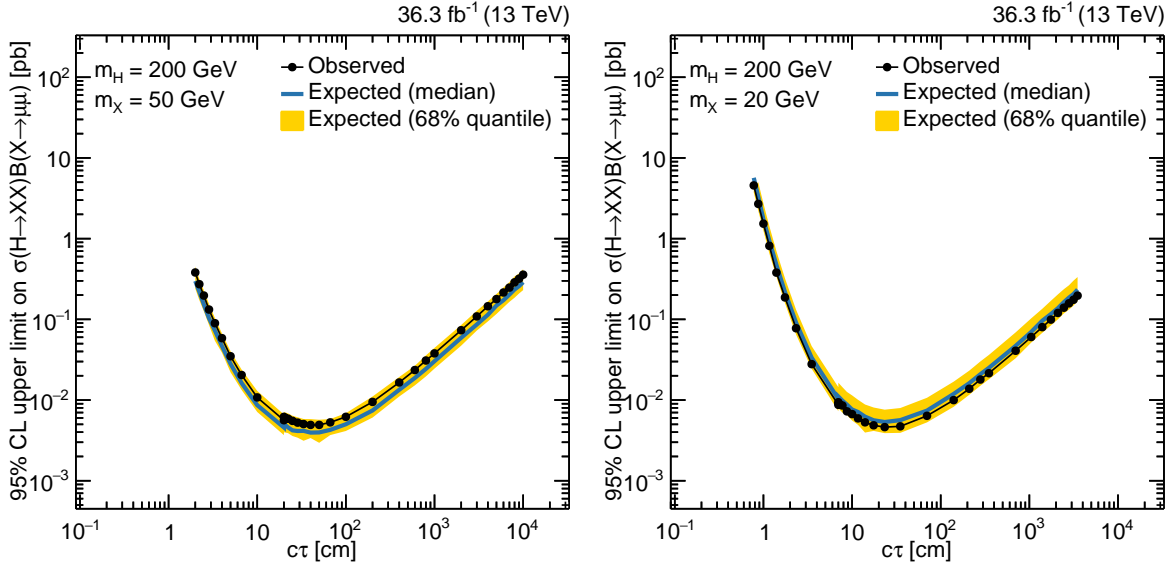


Figure 4.45: 95% CL upper limits on  $\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)$  for  $m_H = 200$  GeV for various values of  $m_X$ . Black dots are observed limits; the blue line represents the median expected limits; the yellow shaded bands show the central 68% quantile.

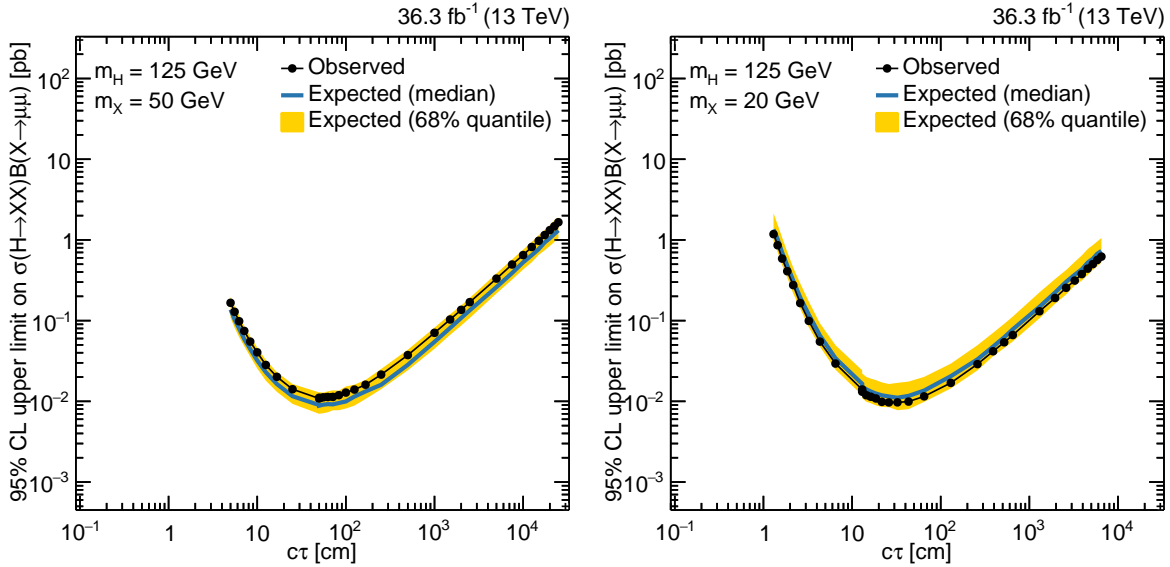


Figure 4.46: 95% CL upper limits on  $\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)$  for  $m_H = 125$  GeV for various values of  $m_X$ . Black dots are observed limits; the blue line represents the median expected limits; the yellow shaded bands show the central 68% quantile.

$m_H$ (GeV)	$m_X$ (GeV)	$c\tau$ (mm)	Expected $[\sigma B]_{UL}$ [pb]	Observed $[\sigma B]_{UL}$ [pb]
1000	350	35	$3.47 \times 10^{-2}$	$3.99 \times 10^{-2}$
		350	$1.69 \times 10^{-3}$	$1.75 \times 10^{-3}$
		3500	$2.98 \times 10^{-3}$	$3.38 \times 10^{-3}$
1000	150	10	$2.53 \times 10^{-1}$	$3.40 \times 10^{-1}$
		100	$2.74 \times 10^{-3}$	$3.67 \times 10^{-3}$
		1000	$1.74 \times 10^{-3}$	$2.35 \times 10^{-3}$
1000	50	4	-	-
		40	$4.24 \times 10^{-3}$	$5.42 \times 10^{-3}$
		400	$1.86 \times 10^{-3}$	$2.32 \times 10^{-3}$
1000	20	2	-	-
		20	$1.12 \times 10^{-2}$	$9.38 \times 10^{-3}$
		200	$6.34 \times 10^{-3}$	$5.40 \times 10^{-3}$
400	150	40	$3.48 \times 10^{-2}$	$4.70 \times 10^{-2}$
		400	$2.04 \times 10^{-3}$	$2.53 \times 10^{-3}$
		4000	$4.72 \times 10^{-3}$	$6.09 \times 10^{-3}$
400	50	8	-	-
		80	$3.44 \times 10^{-3}$	$4.53 \times 10^{-3}$
		800	$2.17 \times 10^{-3}$	$2.50 \times 10^{-3}$
400	20	4	-	-
		40	$7.61 \times 10^{-3}$	$6.51 \times 10^{-3}$
		400	$4.29 \times 10^{-3}$	$3.27 \times 10^{-3}$
200	50	20	$3.90 \times 10^{-1}$	$5.10 \times 10^{-1}$
		200	$4.58 \times 10^{-3}$	$5.62 \times 10^{-3}$
		2000	$7.43 \times 10^{-3}$	$9.48 \times 10^{-3}$
200	20	7	-	-
		70	$1.03 \times 10^{-2}$	$8.73 \times 10^{-3}$
		700	$7.40 \times 10^{-3}$	$6.38 \times 10^{-3}$
125	50	50	$1.05 \times 10^{-1}$	$1.35 \times 10^{-1}$
		500	$9.00 \times 10^{-3}$	$1.08 \times 10^{-2}$
		5000	$2.84 \times 10^{-2}$	$3.74 \times 10^{-2}$
125	20	13	-	-
		130	$1.65 \times 10^{-2}$	$1.40 \times 10^{-2}$
		1300	$2.05 \times 10^{-2}$	$1.69 \times 10^{-2}$

Table 4.12: 95% CL expected and observed upper limits on  $\sigma(H \rightarrow XX)B(X \rightarrow \mu\mu)$  for the 33 sets of signal parameters ( $m_H$ ,  $m_X$ ,  $c\tau$ ) in  $2\mu$  simulated signal samples. Omitted entries refer to signal samples with zero efficiency or with a large statistical uncertainty.

## CHAPTER 5

### Summary and Concluding Remarks

*One sees clearly only with the heart. Anything essential is invisible to the eyes.*

—The Fox, *The Little Prince*, by Antoine de Saint-Exupéry

This thesis presented a search for new long-lived particles decaying to two muons in the CMS detector with pp collision data taken in 2016 at  $\sqrt{s} = 13$  TeV corresponding to  $36.3 \text{ fb}^{-1}$  of integrated luminosity during Run 2 of the LHC. The search presented in this thesis used only the muon system in order to probe the longest lifetimes to which the LHC experiments are sensitive. The results were interpreted in terms of a benchmark model consisting of BSM Higgs bosons decaying to long-lived scalar bosons, but were presented in an inclusive and approximately model-independent way. No excess over the expected Standard Model background was observed.

A notable improvement to the analyses performed with Run 1 data [1, 2] is a higher efficiency for selecting signal events, primarily a result of the development of the DSA to PAT association procedure designed to maintain this efficiency and of the use of the DSA muon reconstruction using cosmic muon seeds over the previously used RSA muon reconstruction. Furthermore, estimating background events with two independent categories—Drell-Yan events and QCD events—allowed the analysis to remain sensitive to long-lived particle masses in a region with significant QCD background. The resulting upper limits are consistent with or improved compared to the analyses performed with Run 1 data.

Along with the results of this thesis work, several potential improvements to the analysis using dimuons formed from two DSA muons have been identified that may be implemented by other CMS researchers in the near future. First, the analysis will benefit from the addition



of data taken at the LHC in 2018, not only from the increased sample size but also from an improved, dedicated displaced dimuon trigger implemented for 2018 data taking. Second, parametrizing certain selections as a function of long-lived particle lifetime, such as the  $L_{xy}/\sigma_{L_{xy}}$  selection, can further discriminate signal events from background events, especially at long lifetimes. Third, technical improvements to the DSA to PAT association procedure involving the segments associated to global-only PAT muons will fine-tune the association and reduce a population of background events. And finally, a technical implementation of the muon timing information for DSA muons, rather than use of the timing from a standalone muon spatially nearby, will further reject out-of-time background events as well as provide an additional method for suppressing cosmic muon events.

In addition, the companion analyses using PAT muons and corresponding dimuons formed from two PAT muons or a PAT muon and a DSA muon will cover the entire sensitivity of the detector. The improved spatial resolution given by the silicon tracker will translate into more efficient signal discrimination and background rejection for long-lived particle decays at a few centimeters.

The exploration of the landscape of new potential long-lived exotic particles is an exciting one, with rapid development in recent years of new analysis techniques to tackle previously unforeseen experimental challenges. This thesis presented a contribution to that exploration, with the hope of illuminating paths to new horizons yet to be undertaken by further studies.

## APPENDIX A

# Neutron-Induced Background Hits in the CMS Endcap Muon System and Implications for the HL-LHC

### Foreword

The work described in this appendix was performed in full collaboration with fellow UCLA graduate student Christian Schnaible under the supervision of our common advisor, Professor Robert Cousins. Together the two students wrote a detailed draft of internal CMS Detector Note DN-2017/019, reproduced below. Professor Cousins also provided editing assistance, particularly in Section A.5. Abhigyan Dasgupta presented this work in a poster session [62] at the 2017 European Physical Society Conference on High-Energy Physics (EPS-HEP), and wrote a conference note for the Proceedings [63]. Christian Schnaible presented this work in a talk at the the 2017 American Physical Society Meeting of the Division of Particles & Fields (APS-DPF) [64].

### Abstract of the Appendix

Among the many challenges to be brought by the high luminosities of the HL-LHC is the impact of increased hit rates in the cathode strip chambers of the CMS endcap muon system. These chambers are used for all levels of trigger as well as offline reconstruction. Neutrons induce background hits via nuclear interactions and capture, followed by gamma emission and (mainly) Compton scatter off electrons that subsequently ionize the chamber gas. This note describes recent efforts to improve the understanding of such neutron-induced background through detailed comparison of CMS pp collision data, GEANT4 simulation, and the response of CMS cathode strip chambers placed in the CERN high-intensity gamma irradiation facility, GIF++. Projections for the effect of such neutron-induced background hits on muon triggering and reconstruction at the HL-LHC are described.

## A.1 Introduction

The high-luminosity upgrade to the CERN LHC (HL-LHC) will bring not only increased access to discovering new physics but also increased background rates in CMS muon chambers [15, 65]. This note focuses on the study of neutron-induced background hits in CMS endcap cathode strip chambers (CSCs) due to their large active volume and forward placement in CMS [34, 66]. Neutrons induce hits in CSCs by producing photons, via either nuclear scattering or capture. These photons produce or scatter electrons, via Compton scattering, the photoelectric effect, or pair production; the ionization from these electrons can either corrupt hits from desirable tracks in the event, or add extra background hits. This note describes efforts to quantify and characterize the effects of this neutron-induced ionization through analysis of Monte Carlo (MC) simulation of neutron production and propagation in CMS, as well as of CMS proton-proton (pp) collision data. The results are tied to analyses of muon test beam data at the CERN Gamma Irradiation Facility (GIF++), which allows us to study the effects of background radiation on CSC performance and project the effects of neutron-induced background to the conditions at the HL-LHC.

In the debris of pp collisions, hadronic interactions liberate neutrons from CMS calorimeters, shielding, and various other detector structures. These neutrons can carry several GeV or more of kinetic energy, propagating throughout the experimental cavern, detector materials, and shielding. They lose kinetic energy through many scattering interactions over the course of several milliseconds, and if not captured sooner, are eventually cooled to thermal equilibrium with the cavern environment, carrying kinetic energies of around 0.025 eV (300 K). There are three main processes by which neutrons lead to photons: inelastic scattering, resonant capture on nuclei, and thermal capture on nuclei [67].

Neutrons with several MeV of kinetic energy may scatter inelastically with detector and shielding material, producing one or more photons. In addition to undergoing inelastic scattering, neutrons may be captured by various nuclei, including iron, copper, and free hydrogen, resulting in excited isotopes of the original capturing nuclei with energies typically a few MeV greater than that of their ground states. The excited isotopes reach their ground

states by emitting one or more photons. Resonant capture occurs when the kinetic energy of a neutron (typically of order keV) is such that the sum of the total neutron and target nucleus energies matches the total energy of a discrete excited state of the final-state isotope. For example,  $^{56}\text{Fe}$  has a resonant neutron capture cross section at a neutron kinetic energy of 1.1 keV [67]. Thermal capture occurs when neutrons have reached sub-eV thermal energies, in a regime where the cross section for capture increases with decreasing neutron velocity. Throughout this note, any reference to neutron capture refers to both thermal and resonant neutron capture unless otherwise specified.

The photons produced from inelastic scattering and nuclear capture are typically of order MeV in energy. Once produced, the photons can propagate inside the gas volume of a CMS CSC where they predominantly scatter off electrons, but also produce electrons via pair production and the photoelectric effect. These scattered electrons can ionize the chamber gas and subsequently lead to background hits.

Such hits constitute an important source of background rates as well as contribute to aging of muon chambers. This note focuses on the effect that these increased background rates have on muon triggering and offline reconstruction of muons and measurement of their properties.

The identification of neutron-induced hits in CMS data is complicated by the fact that data from a CSC is read out only if the CSC trigger electronics identify a potential muon track (trigger primitive) in that CSC [68, 69, 70]. A muon leading to a trigger primitive can induce other hits in the CSC via e.g. bremsstrahlung or delta rays, or be accompanied by other charged particles from a jet in which the muon is embedded. Precision measurements of the ionization charge are read out through the main CMS data acquisition (DAQ) system only for a small fraction of the area of the CSC that is near the muon that generated the trigger, and are particularly vulnerable to hits from these extra particles. Thus, for most of the studies in this paper, we use instead the diagnostic information for the whole chamber that is passed to the DAQ by the trigger electronics (the so-called “trigger path”). These trigger path data are much less precise than the high-precision charge data read out by the main DAQ [71], but since they include information from the entire chamber, they allow us to

look for neutron-induced hits in areas of the chamber that are further away from the muon that generated the trigger. With various additional selection criteria, sufficient samples of neutron-induced hits are obtained for detailed studies of rates in CMS CSCs, for comparison to simulated CMS data and GIF++ data.

Section A.2 is a brief overview of cathode strip chambers and the on-chamber trigger electronics in the CMS detector. Section A.3 describes GEANT4 simulation of the detector response to pp collisions, including neutron-induced hits. Section A.4 describes the selection criteria that we use to isolate a sample of neutron-induced hits in CMS CSC data. Section A.5 describes the procedure by which we normalize our neutron-induced hit counts into hit rates. Sections A.6 and A.7 describe studies of GIF++ data and CMS CSC data from which we predict some effects of increased radiation in HL-LHC-like conditions.

## A.2 Cathode Strip Chambers and Trigger Electronics in CMS

The CSC system in the endcap muon detection system of CMS [69] contains 540 chambers. Each endcap has 4 stations numbered from 1 to 4, and each station contains two or three rings of chambers. A chamber is uniquely specified by a string of the form  $\text{ME}\pm S/R/C$ , where  $\pm$  refers to either the + or – endcap,  $S$  is the station number,  $R$  is the ring number, and  $C$  is the chamber number, e.g. ME+1/1/36. Part of this notation may be omitted when referring to groups of chambers; for example, we refer to ME1/1 or ME2/1 chambers, i.e. chambers in the innermost ring of the first or second stations, respectively, in either endcap.

The ring of chambers in each station is perpendicular to the  $z$  axis of CMS, with the narrow end of their trapezoidal shapes closer to the beam line. Figure A.1 contains a diagram illustrating the principle of operation of a CSC. Each CSC is a multi-wire proportional chamber consisting of 6 layers, with each layer lying in an  $r$ - $\phi$  plane of CMS. Each layer contains a gas volume consisting of 50% CO<sub>2</sub>, 40% Ar, and 10% CF<sub>4</sub> in between a plane of copper cathode strips and a plane of anode wires. The strips extend from the narrow end to the wide end, along the  $r$  direction, at constant  $z$  and  $\phi$  coordinates; the wires extend parallel to the narrow and wide ends, approximately along the  $\phi$  direction, at constant  $z$  and

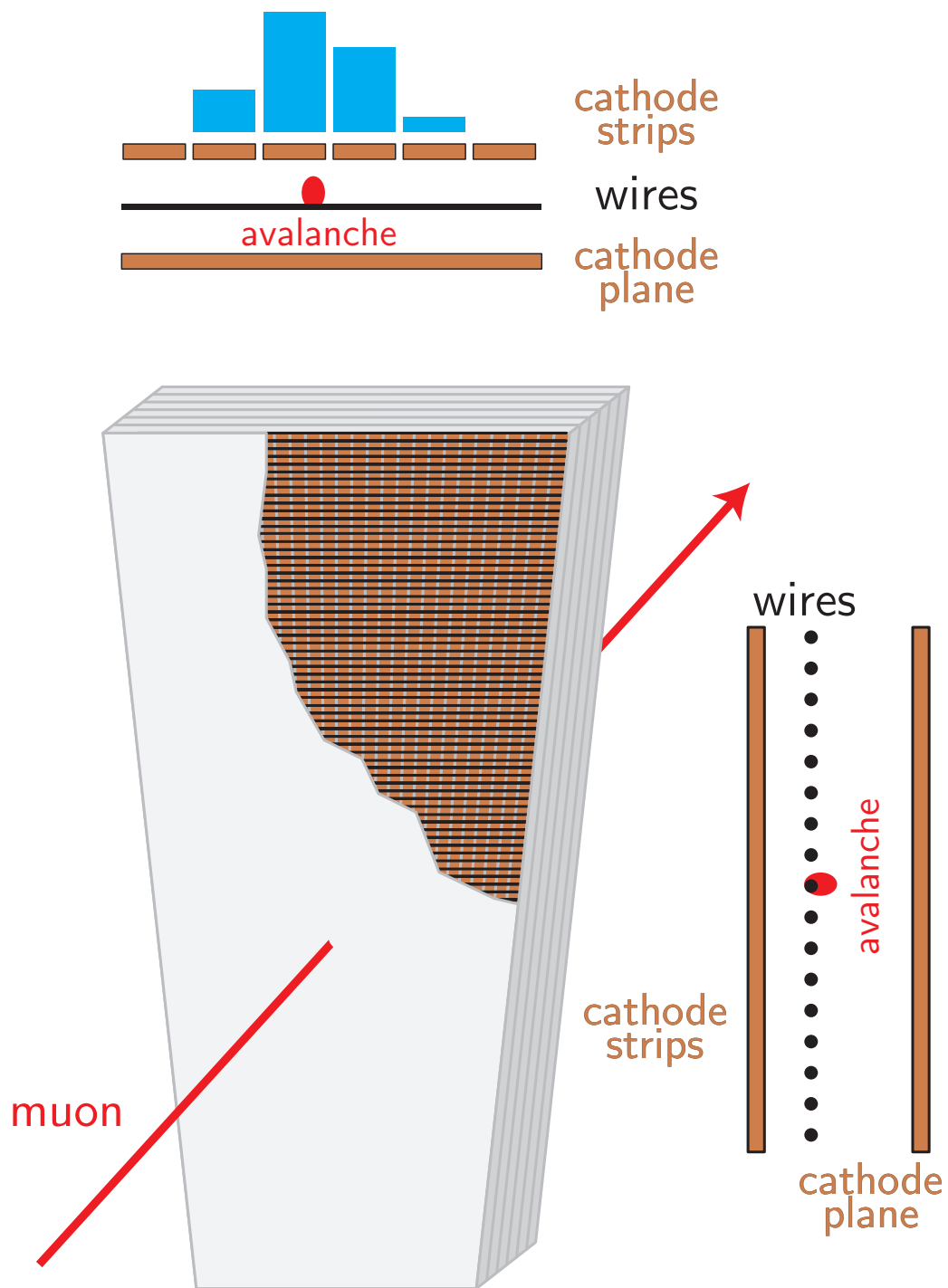


Figure A.1: Diagram and principle of operation of a CSC endcap muon chamber in CMS.

approximately constant  $r$  coordinates. The wires are spaced 2.5–3.5 mm apart, arranged in groups of 5–16 mm (wire groups, or WG), and kept at a high voltage of 2900–3600 V with respect to the cathode strips.

A muon ionizes the gas, and the resulting electrons drift towards the wires, causing an avalanche of charge that induces an opposite charge on the cathode strips. The notable feature of a CSC is excellent position resolution in the  $\phi$  direction achieved by precision cathode charge readout and interpolation. Cathode trigger electronics include a comparator network, which provides half-strip hit position resolution in the trigger hardware. Cathode comparator hits give the layer and half-strip number and anode wire group hits give the layer and wire group number. A digitized detector hit – a cathode strip hit or a wire group hit – is referred to as a digi, and therefore the strip (or half-strip) number or wire group number may also generally be referred to as a digi number [68, 71, 70].

As illustrated in Figure A.2, the anode wire group hits and comparator half-strip hits are transmitted to circuits that perform low-level local pattern recognition on the hits in the six layers of a chamber. Candidate tracks from charged particles are identified separately in the cathode strips and anode wires and then combined. Anode local charged tracks (ALCTs) are formed by ALCT boards, which receive data for each plane from anode front end boards (AFEBs). Cathode local charged tracks (CLCTs) are formed in two steps within the trigger motherboard (TMB) from data transmitted by cathode front end boards (CFEBs). First, a set of loose requirements, for example on the number of layers hit, are applied to the CFEB data. These are known as the pre-trigger, and a candidate track that passes these requirements is called a pre-CLCT. Then, a set of tighter requirements are applied, and a pre-CLCT that passes these requirements becomes a CLCT. The TMB also receives ALCTs from the ALCT boards and combines coincident ALCTs and CLCTs to create local charged tracks (LCTs). The CSC electronics firmware has programmable requirements for the various pattern recognition steps, e.g. for the number of layers hit, the spatial distribution of the hits, the temporal distribution of the hits, and the time window for coincidence of the ALCTs and CLCTs.

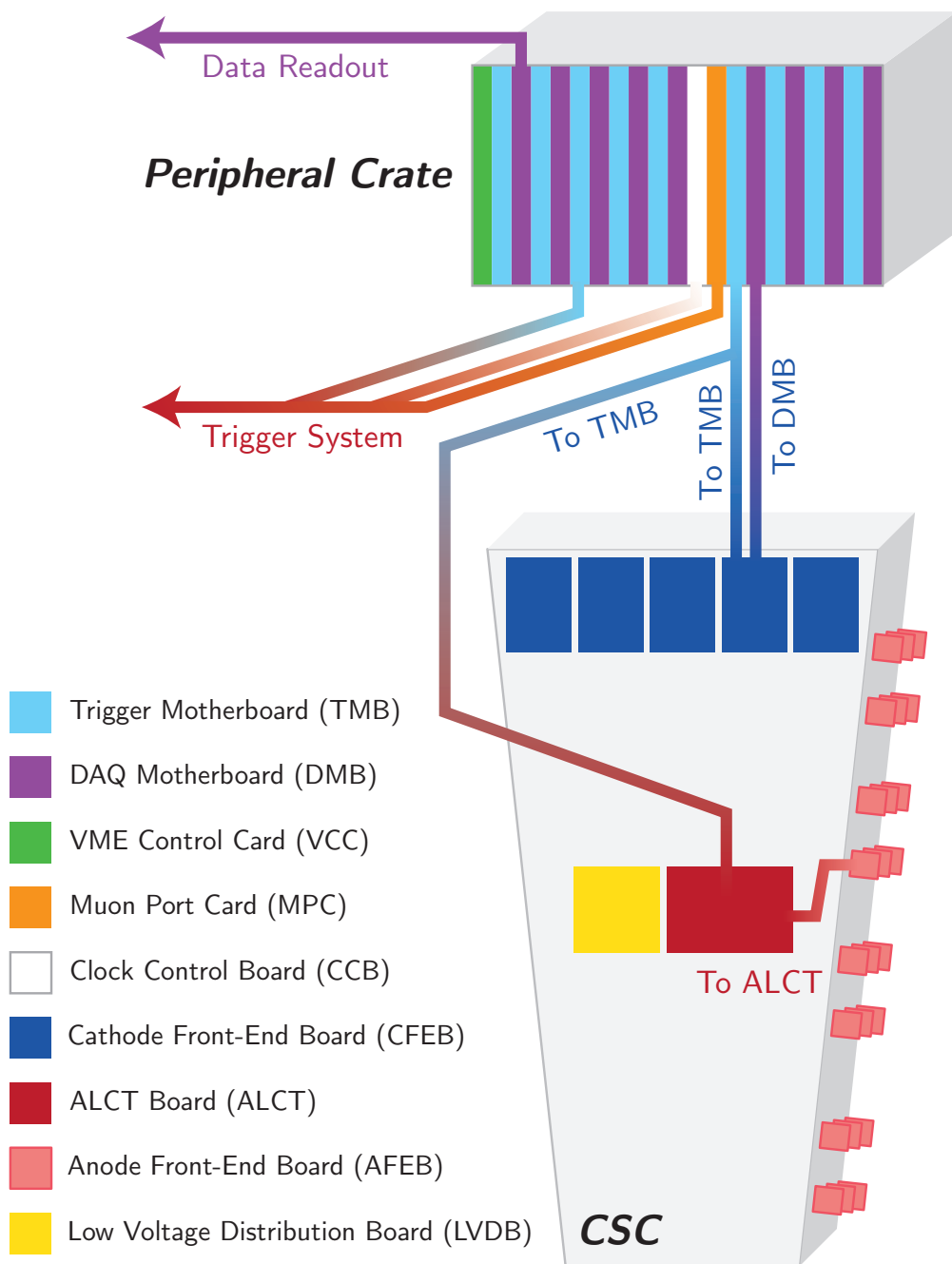


Figure A.2: Diagram of the CSC electronics system in CMS.



These trigger data, referred to as trigger primitives, are transmitted to the endcap muon trigger system and also inserted into the DAQ stream via the DAQ motherboard (DMB). For offline use, the DMB also reads out the high-precision analog-to-digital converter (ADC) charge measurement for each cathode strip when the DMB receives a signal from the CMS Level-1 trigger. However, these ADC charge measurement data are read only from CFEBs for which a pre-CLCT was formed, which correspond to only a fraction of a chamber. As noted above, muon-induced background hits make these data of limited use in our current studies of neutron-induced hits.

The data from the TMB contains both early and late detector hits with respect to the muon that generated the trigger are inserted into the DAQ stream. Hit times are binned into 16 readout time bins of 25 ns each, with the muon usually inducing hits in time bins 7 or 8. The chamber electronics can thus read out detector hits that occur approximately 200 ns (8 time bins) before and after the muon hits.

### A.3 Geant4 MC Simulation for Neutron Studies

#### Geant4 MC Simulation Setup

To better understand the basic nuclear interactions that may result in background hits in CSCs, we turn to simulation. We used the GEANT4 simulation package to simulate minimum-bias proton-proton (pp) collision events [45, 46, 47, 72]. To study neutron background effectively, a few specific modifications to the default CMS simulation configurations were necessary [73];

- To accommodate the long lifetimes of the neutrons under study, we extended the tracking time of all particles to 10 s.
- To accommodate the low energies of the neutrons under study, we removed any particle energy thresholds where they existed.

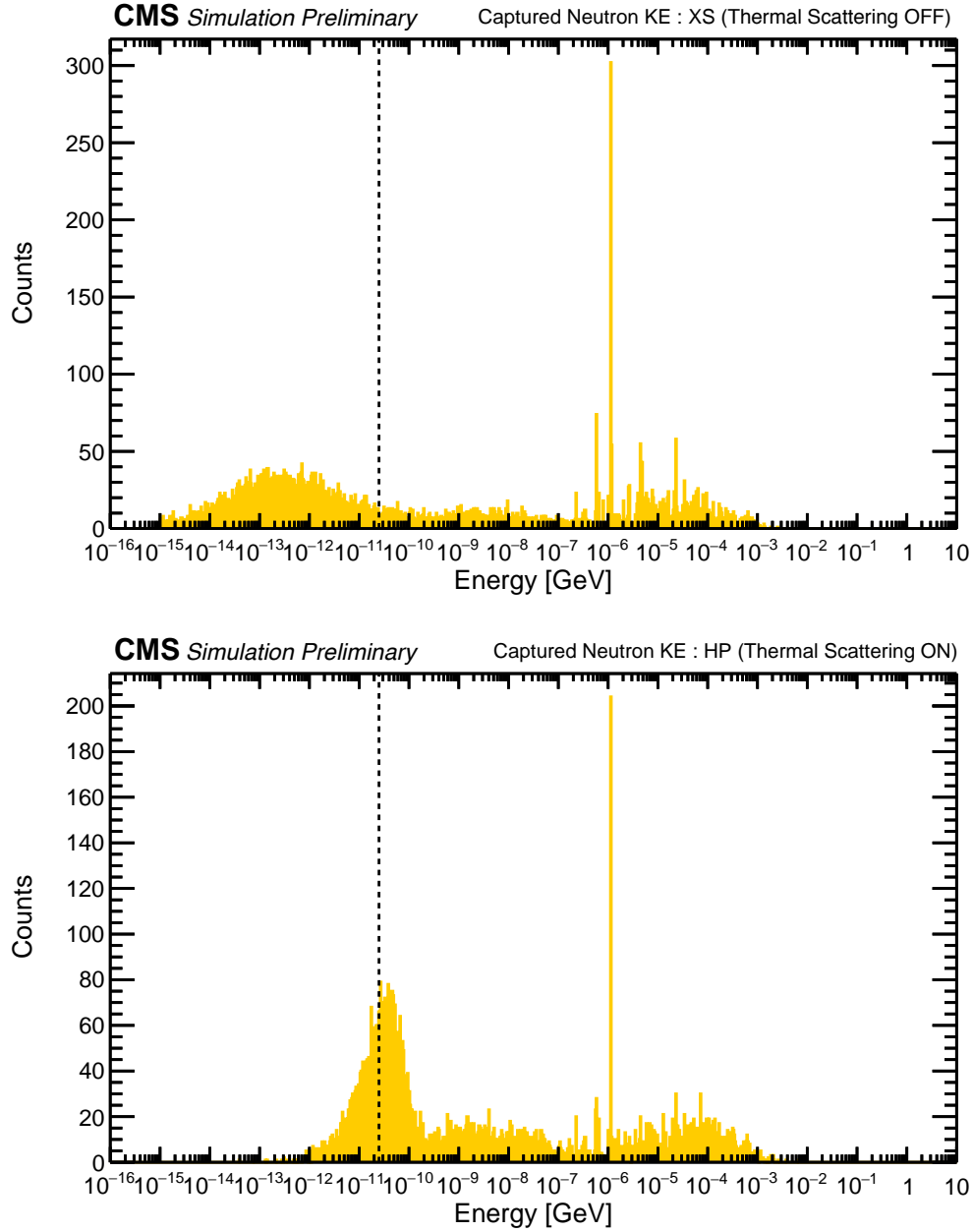


Figure A.3: Histogram of neutron kinetic energy just before nuclear capture, (top) before and (bottom) after enabling thermal treatment of detector nuclei. The dashed line indicates 0.025 eV. Before enabling the new feature, the low energy peak of captured neutrons is several orders of magnitude below thermal energies; after enabling the feature, the low energy peak is at thermal energies. Also visible are the resonant capture peaks on, for example,  $^{56}\text{Fe}$ , in the keV range.

- To properly model low-energy neutrons, we enabled a feature in the GEANT4 thermal neutron scattering routine that models nuclei at room temperature. The standard GEANT4 routine models the temperature of the nuclei in the detector at 0 K and as a result artificially cools neutrons down to well below thermal energies. Figure A.3 displays histograms of the kinetic energy of neutrons just before their capture on a nucleus, before and after the new feature was enabled. The dashed line is at 0.025 eV; without the feature enabled, the low energy peak of captured neutrons is well below thermal energies, whereas with it enabled, the low energy peak is at thermal energies.
- To retain all electronics signals in simulated CSCs, we assigned simulated hits produced after times longer than 200 ns a time of 200 ns in the CSC digitization simulation module. This is to ensure that all hits from a single simulated pp collision are retained, rather than just the hits that occur within the simulated detectors' limited time readout window of 400 ns.

Two versions of GEANT4 neutron interaction packages are compared:

- HP: “High Precision” package that parametrizes existing experimental nuclear data for neutron interaction cross sections, and
- XS: intended for CPU time optimization, derived from HP by approximating detailed parametrizations in HP with averages.

## Results from MC Simulation

In examining these simulated events, we find that each pp collision can lead to three broad categories of neutron-induced hits:

- Hits on the time scale of 100 ns after the pp collision, induced by “fast” neutrons whose energies have degraded over time to a few MeV. The neutrons interact inelastically with nuclei, resulting in some ionization from protons or nuclear fragments, but primarily resulting in nuclear de-excitation photons that give rise to ionizing electrons, primarily from Compton scattering.

- Hits on the time scale of several  $\mu\text{s}$  after the pp collision, induced by neutrons whose energies have degraded over time to energies of a few keV. The neutrons are resonantly captured on various types of nuclei, also resulting in nuclear de-excitation photons that give rise to ionizing electrons.
- Hits on the time scale of several ms or longer after the pp collision, induced by thermal neutrons whose energies have degraded over time to be at thermal equilibrium with the experimental cavern. The neutrons are captured on nuclei, also resulting in nuclear de-excitation photons that give rise to ionizing electrons.

Figure A.4 displays plots of the final energy of simulated neutrons vs. the time (since the pp collision) of the corresponding simulated detector hits (SimHits) in CSCs. The final energy of the neutron refers to the energy of the neutron at the moment it gives rise to the photon that eventually gives rise to the electron that produced the SimHit. Visible on the top plot are the three categories of hits as mentioned above: two groups of red dots that are hits induced by neutron captures; and blue dots that are hits induced by neutron inelastic scatters. The bottom plot in Figure A.4 shows the same quantities for all neutron-related SimHits, including the photon-induced hits as well as hits from ions (green) and protons (magenta).

Since the time scales from when the neutrons are created to when neutron capture induces hits are much longer than time scales of LHC bunch train gaps [15], hits occur in chambers uniformly at all times during an LHC fill. In simulation, we distinguish thermal neutron capture-induced hits in CSCs from prompt and fast neutron-induced hits by selecting the simulated hits that occur after a sufficiently long time after the pp collision.

Figure A.5 is an  $r$ - $z$  view of the CMS cavern showing the location of neutron captures that lead to SimHits in the CSCs, each dot colored according to the capturing nucleus. Neutrons are mainly captured on  $^{56}\text{Fe}$  (red),  $^{63}\text{Cu}$  (green), and  $^1\text{H}$  (blue) within the CMS endcap muon system, CMS calorimeters, and LHC shielding.

Of additional interest are the processes by which photons produce the electrons that result in SimHits. Figure A.6 is a histogram of the simulated photon energy emitted by nuclear de-

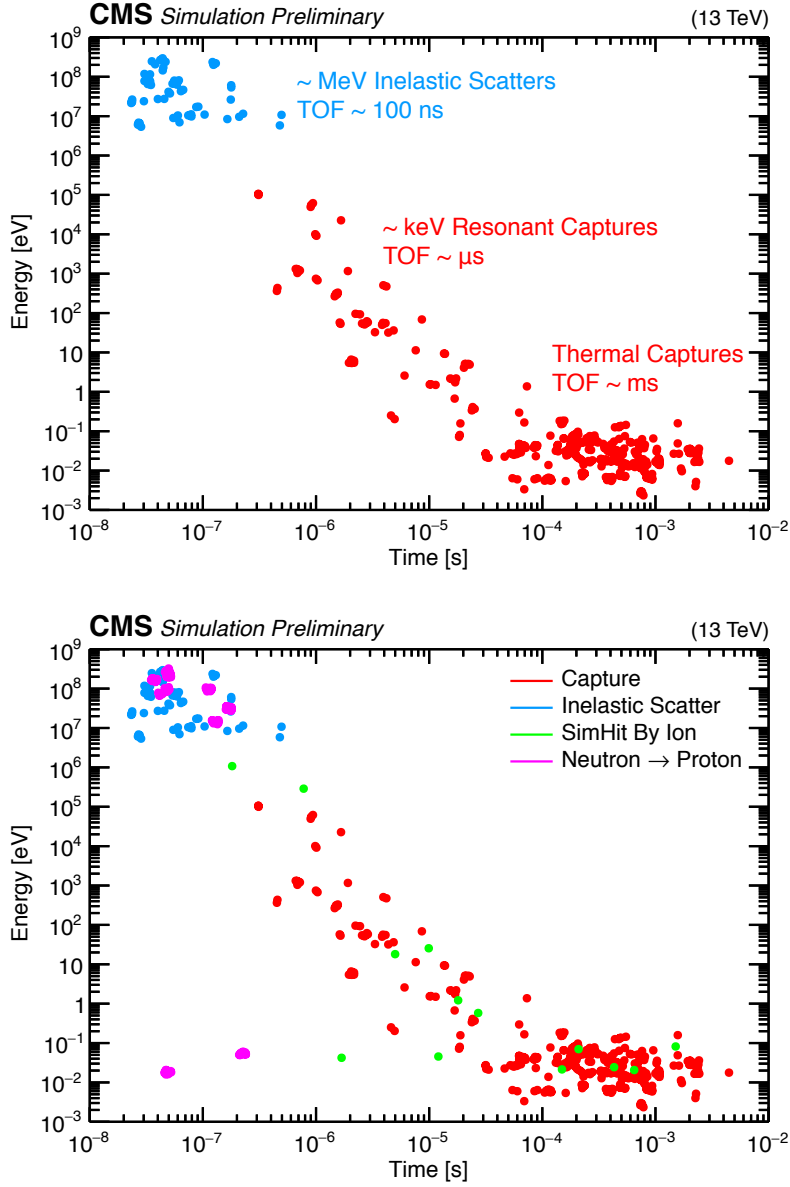


Figure A.4: Final energy of simulated neutron vs. the time since pp collision giving rise to simulated detector hit, for hits in CSCs. Hits are induced by electrons that are produced from photons that are produced from neutron capture or from neutron inelastic scattering. Red dots indicate simulated hits induced by neutron captures, and blue dots indicate simulated hits induced by neutron inelastic scatters. The top plot shows hits from photons only, coming from inelastic scattering and nuclear capture; the bottom plot shows all neutron-related hits, including those induced by protons and nuclear fragments, shown in magenta and green dots, respectively.

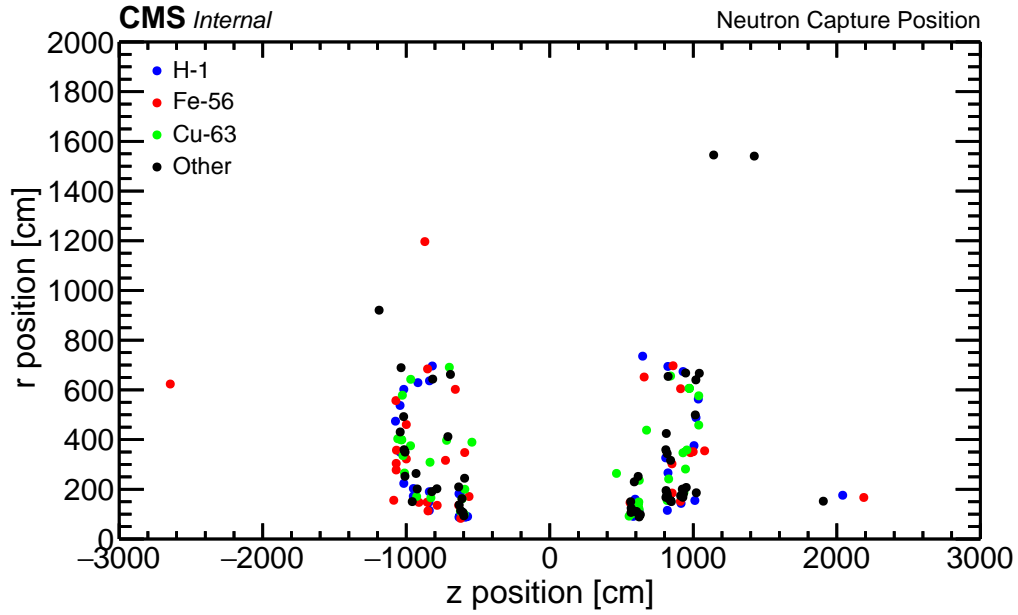


Figure A.5:  $r$ - $z$  view of CMS cavern showing the locations of the specific neutron captures that led to SimHits in the CSCs.

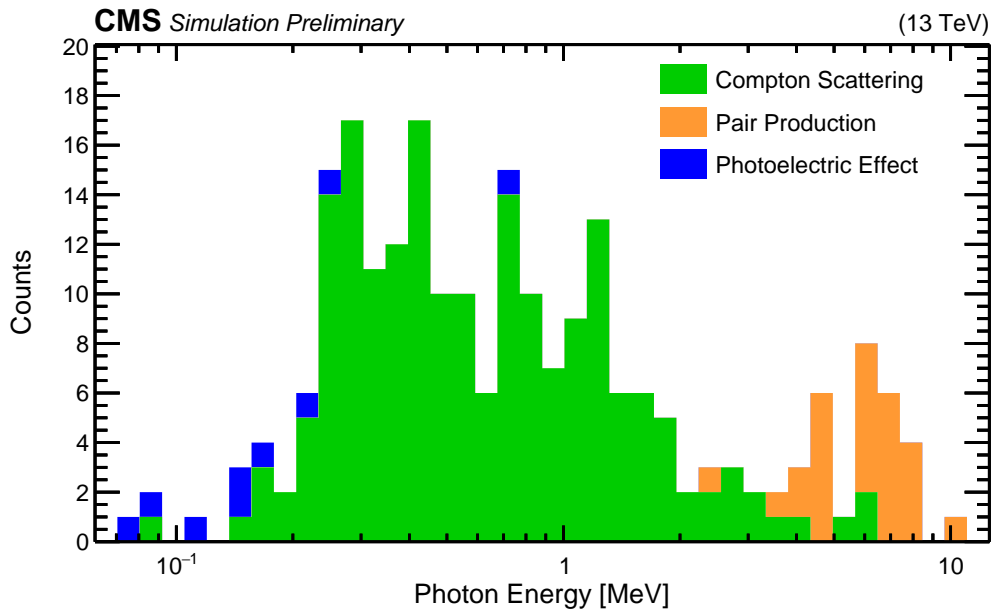


Figure A.6: Stacked histogram of the final energy of simulated photons at the time they produce simulated electrons that lead to simulated detector hits. The photons are categorized by the process by which the electrons are created or scattered. The most common process by which electrons that eventually lead to simulated hits are formed is Compton scattering.

excitations from both inelastic scattering and neutron capture that lead to SimHits, stacked by the specific production or scattering process that created the ionizing electron. Photons that lead to SimHits predominantly have energies of hundreds of keV to several MeV, and the most common process that leads to SimHits is Compton scattering.

## A.4 Selection and Isolation of Neutron-Induced Hits in CMS Data

To obtain a sample of events in pp collision data suitable for studying the neutron background, we first select events with Z boson candidates decaying to two opposite sign muons from the SingleMuon dataset from Run 2016H, corresponding to approximately  $9\text{ fb}^{-1}$  of integrated luminosity. Events are chosen with at least two muons of opposite sign where the muon leading in  $p_T$  is required to have  $p_T > 30\text{ GeV}$ , and the subleading muon is required to have  $p_T > 20\text{ GeV}$ . In addition, muon candidate tracks are required to have at least one hit in the pixel detector, hits in at least six silicon-strip tracker layers, and segments in two or more muon stations. To suppress non-prompt muons, the sum of the  $p_T$  of charged tracks within a cone of  $\Delta R = 0.3$  around the muon is required to be less than 10% of the muon  $p_T$ . Also, the muon is required to have a transverse impact parameter of less than 2 mm and a longitudinal distance of less than 5 mm with respect to the primary vertex. The Z boson dimuon candidates are selected within an invariant mass window of 60–120 GeV. We thus obtain a sample of prompt muons with reduced hadronic activity near the muon candidate. Finally, at least one of the muons is required to have passed through at least one CSC, by requiring that an LCT in at least one CSC was matched to the muon candidate track.

We then look for evidence of neutron-induced hits in cathode comparator half-strip hits and anode wire group hits in the data of each CSC thus selected. To reduce potential muon-induced contamination of our selection of candidate neutron-induced hits, we require that there be exactly one LCT occurring in a one-sixteenth corner of a chamber and consider only comparator half-strip or wire group hits occurring in the opposite half of the chamber from the LCT. Figure A.7 displays an example diagram of a CSC with a muon (indicated by the black arrow) passing through a chamber corner (shaded in red) and the corresponding

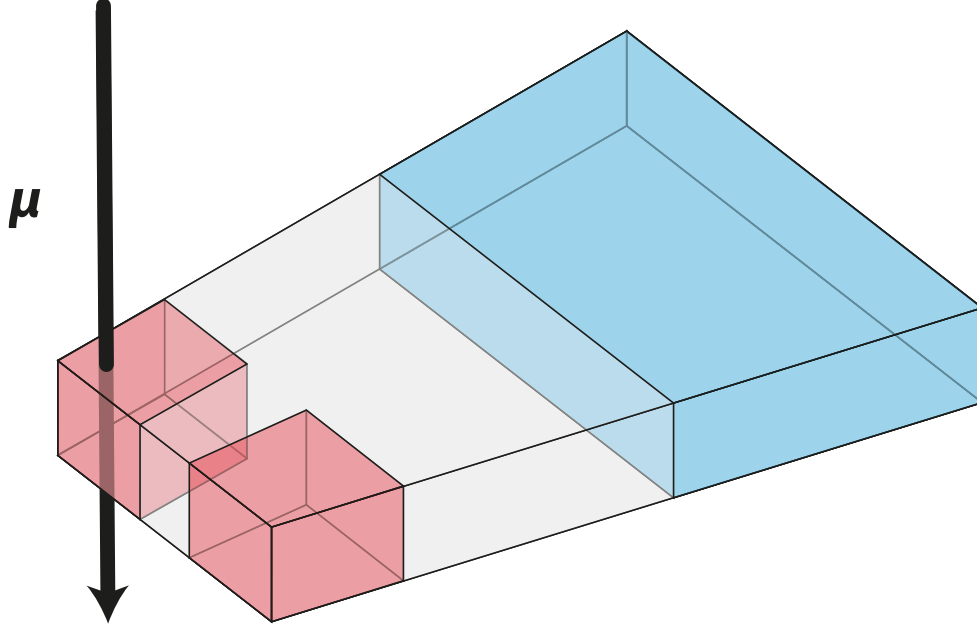


Figure A.7: Opposite half selection for neutron-induced hits in CSCs.

opposite half of the chamber (shaded in blue) where we look for hits. This ensures that any potential muon-induced background is at least a quarter chamber spatially separated from the region of the chamber in which we search for candidate neutron-induced hits.

As a precise clarification of what is meant by a one-sixteenth corner and an opposite half, an example definition of a chamber corner area is the overlap region between wire groups 1 through  $N_{\text{WG}}/4$  and comparator half-strips 1 through  $N_{\text{HS}}/4$ , where  $N_{\text{WG}}$  and  $N_{\text{HS}}$  are the total number of anode wire groups and comparator half-strips in a chamber, respectively. The corresponding area for the opposite half of the chamber in which we search for candidate neutron-induced wire group hits is the area defined by wire groups  $N_{\text{WG}}/2$  through  $N_{\text{WG}}$ , and similarly for candidate neutron-induced comparator half-strip hits is the area defined by  $N_{\text{HS}}/2$  through  $N_{\text{HS}}$ . All four definitions of a chamber corner and their corresponding opposite halves are considered and used in this analysis.

Since the rate of trigger muons varies over the CSC system, the number of times we look in each half-chamber varies for each half-chamber in the CSCs and for each event, and the



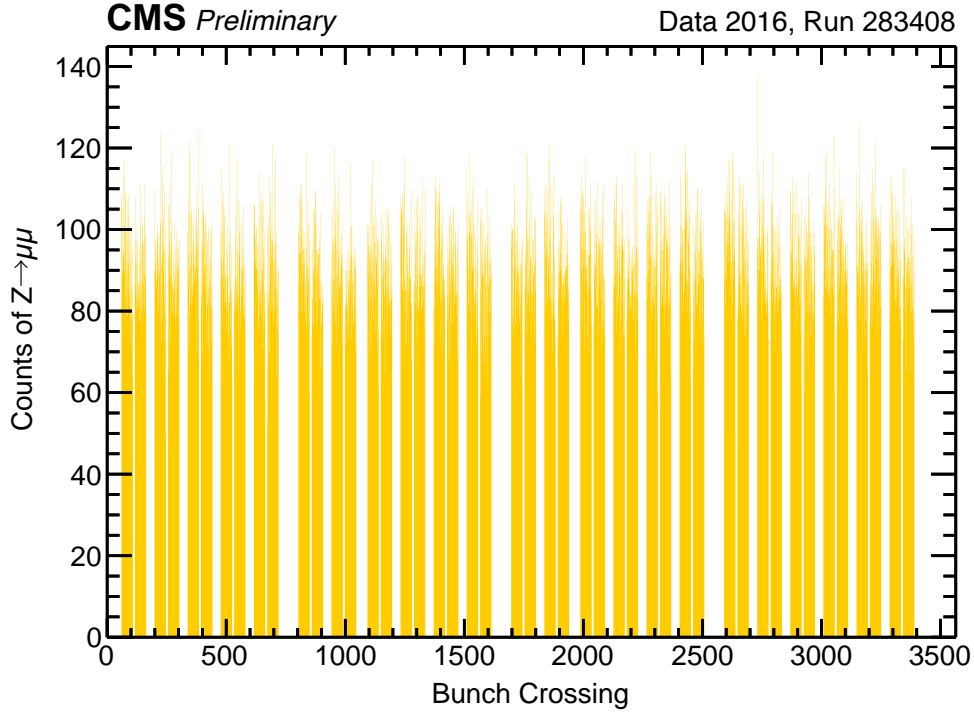


Figure A.8: Histogram of bunch crossing number in a particular CMS data taking period during LHC proton-proton collisions in 2016, showing the gaps of various sizes in the LHC bunch structure.

hit rates vary with luminosity. We introduce some rather tedious bookkeeping to compute rates in the next section, with an indicator function used to specify when we look.

The LHC proton beam has 3564 bunch places, with a bunch spacing of 25 ns. Not every bunch place has protons in it; rather, consecutive bunches of protons occur in trains, separated by gaps [15]. Proton-proton collisions occur in CMS when bunches of protons cross, called bunch crossings (BX). Figure A.8 is a histogram of LHC bunch place numbers for a high statistics run. The non-empty histogram bins correspond to bunch places that are filled with pp collisions, displaying the LHC bunch structure. In this run, trains with protons are exactly 48 bunches long and the trains are separated by bunch gaps of 8 to 35 bunch places, with the final train followed by the LHC abort gap of approximately 200 bunch places.

In each event read out by the CMS DAQ, anode wire hits are read out in 16 time bins of 25 ns. As mentioned in Section A.2, hits in time with the muon occur in the middle of the readout window in time bin 8; time bins before time bin 8 then correspond to earlier bunch places. Hits from these early-time time bins where the readout was triggered by a muon from pp collisions in the first BX after a gap correspond to times within the gap. That is, these hits occur during bunch places without any pp collisions. Thus, they must arise from pp collisions from before the start of the gap (apart from the rare cosmic or noise hit), and we attribute them to be neutron-induced hits from long-surviving neutrons, i.e. from neutron capture. Events with muons that are triggered in the first few BXs after a gap can also be used to identify neutron-induced hits as long as the early-time time bins are early enough to occur within the gap.

Similarly, events (triggered by muons) at the end of a bunch train have late-time time bins that correspond to times within the beginning of the next bunch gap. As the LHC bunch trains often have 48 consecutive BX with collisions between gaps, for convenience we consider only bunch trains that are exactly 48 BX long and that occur after gaps of at least 35 bunch places.

Figure A.9 is a 2D histogram of the CSC anode wire hits (intensity given by color code) as a function of the number of bunch crossings after a gap on the  $y$  axis and the digi readout time bin on the  $x$  axis. Both BX and time bins are 25 ns wide. Time bin 8 is in-time with the muon that triggered readout of the chamber.

The region bounded by the lower left red outlined triangle contains early readout time hits in early bunch places, i.e. hits recorded near the end of an LHC bunch gap. This triangle contains neutron capture induced hits with negligible contribution from cosmic muons. The excess hits in readout time bin 0 contain hits from the previous bunch crossing due to the length of electronic pulses and are therefore not considered. We use the bins selected within the lower left triangle as the final step in isolating neutron capture induced hits.

The central red outlined rectangle is a region delineating early readout time hits in bunch places contained completely within a bunch train. Hits from these bunch crossings consist of,

in addition to neutron capture induced hits, hits induced by fast neutrons and hits in time with pp collisions from other bunch crossings occurring in out-of-time readout time bins. The upper right red outlined triangle is a region delineating hits in late bunch places, i.e. hits recorded at the beginning of an LHC bunch gap. Hits from these bunch crossings contain not only neutron capture induced hits, but also hits induced by fast neutrons (neutrons which have not yet lost enough energy to be captured) that occur within a few bunch crossings of a pp collision. Table A.1 summarizes the time windows for each of the three regions outlined in Figure A.9.

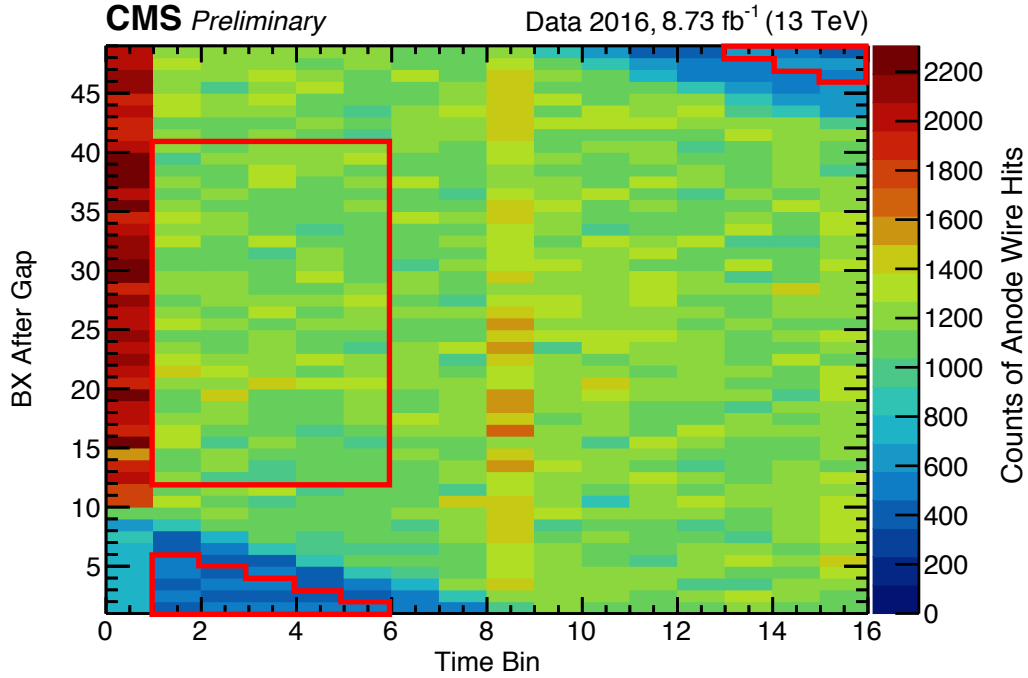


Figure A.9: 2D histogram of number of anode wire hits from CSCs in the inner ring of the first station (ME1/1) of the CMS endcap muon system. The histogram is plotted as a function of the readout time bin, and of number of bunch crossings (BX) after an LHC bunch gap of at least 35 BX in bunch trains that are exactly 48 BX in length. Hits in readout time bin 0 are contaminated with an artifact of anode wire readout electronics and are therefore not considered.

Table A.1: Enumeration of time windows used for each region in Figure A.9, listed by digit readout time bins and bunch places after the gap.

Region	BX After Gap ( $b$ )	Readout Time Bins ( $T(b)$ )	$N_{T(b)}$
Lower left triangle (neutron capture only)	1	1–5	5
	2	1–4	4
	3	1–3	3
	4	1–2	2
	5	1	1
<b>Total 2D bins in region</b>			<b>15</b>
Central rectangle (neutron capture, fast neutrons, and hits from pp collisions)	12–40	1–5	5
<b>Total 2D bins in region</b>			<b>145</b>
Upper right triangle (neutron capture, fast neutrons)	46	15	1
	47	14–15	2
	48	13–15	3
<b>Total 2D bins in region</b>			<b>6</b>

In summary, we select neutron capture induced hits in CMS data using the following criteria:

- Triggering muons must be from  $Z \rightarrow \mu\mu$  candidates in SingleMuon Run 2016H
- Exactly one LCT must be in a one-sixteenth corner of a chamber
- Digis must be in the opposite half of the chamber from the LCT
- Digis must be found at early times in events occurring in the first few bunch places in a train of size exactly 48 bunch places, after a bunch gap of at least 35 bunch places

## A.5 Results and Comparison of CMS Data with MC Simulation

### Computation of Neutron-Induced Hit Rates

With the selection of neutron-induced hits in CMS data as well as the understanding gained from neutron simulation at CMS, we examine the local  $r$  and  $\phi$  distributions of neutron-induced hits within a given chamber type, as well as the neutron-induced hit rate for a given chamber as a function of instantaneous luminosity [74, 54]. Directly comparing CMS data and MC simulation requires careful normalization of the digi counts. This section describes how the samples of neutron-induced hits obtained using the procedures discussed in Section A.4 are normalized, and various neutron-induced hit rates are defined.

As discussed in Section A.4, neutron captures typically occur many bunch places after the pp collisions that created the neutrons, so that, to a good approximation, neutron-induced hits uniformly populate all bunch places in the LHC. Observation of early-time neutron-induced hits during a particular triggering bunch crossing necessarily implies contributions from many pp collisions in many previous bunch crossings at CMS. In the steady state of approximately constant instantaneous luminosity, we can associate the aggregate number of all neutron-induced hits to the aggregate number of all pp collisions, even though the pp collisions are not uniformly spaced in time at CMS, having significant gaps between bunch trains. This association is the basis for all our rate calculations.

Since we connect rates of neutron-induced hits to approximately constant instantaneous luminosity, it is convenient to use the instantaneous luminosity as measured over a “lumi section,” a time interval of data taking at CMS of about 23 s. We consider any changes in instantaneous luminosity within a lumi section to be negligible. We further consider edge effects from neutrons created during one lumi section but detected during another to be negligible. (The effects on the two ends of the a lumi section tend to cancel in any case.)

As further discussed in Section A.4, the neutron-induced hits that we observe are a random sample of the steady stream of hits, where the number of samples varies from half-chamber to half-chamber within CMS and with each event, according to the selection criteria described. Thus, in order to normalize our neutron-induced hit rates, we keep track of the number of times we look for hits in each half chamber as well as the steady state rate of pp collisions at the times we look.

The bookkeeping accounts for each 2D bin in Figure A.9, each readout time bin and bunch place after the gap. The details are described in the next sections; here we define some variables to be used.

Each ring specifies a different chamber type, represented by  $c$ . These chamber types contain different numbers and sizes of wire groups and half-strips; we let  $d$  (for digi) represent a particular wire group or half-strip within a given chamber type. The counts also vary with the particular bunch place  $b$  after a gap and the particular readout time bin  $\tau$  in which we look.

## Computation of Neutron Hit Rate in CMS Data

To determine the neutron-induced hit rate in CMS data, we count the number of selected digis found in particular bunch places after a gap and particular time bins in one of the red outlined regions in Figure A.9. These regions consist of a set  $B$  of bunch places after a gap, and a set  $T(b)$  of time bins for each particular bunch place  $b \in B$  after a gap. Table A.1 enumerates each of the bunch places after a gap considered in each region, consisting of a bunch place set  $B$ , the set of time bins  $T(b)$  for each bunch place, and the total number of

time bins  $N_{T(b)}$  for each bunch place. For all  $b$  not in the table,  $T(b)$  is the empty set.

We process our data as a series of DAQ events, as described in Section A.4. In the following, “event” refers to a DAQ event, which corresponds to a trigger at a particular bunch crossing leading to a complete readout of the detector. The event data of course contains hits from many pp collisions, including (in the out-of-time time bins) hits from pp collisions in other bunch crossings.

We formalize our notation for counting the number of neutron-induced digis in a somewhat tedious manner in order to help with the bookkeeping that follows for keeping track of the associated pp collisions.

Let  $N_{\text{events}}^{\text{CMS}}$  be the number of events in the data set used; we let the index  $i$  run over the events considered, so that  $i = 1, \dots, N_{\text{events}}^{\text{CMS}}$ .

Let  $b(i)$  be the number of the bunch place after a gap for event  $i$ .

Then  $T(b(i))$  is the set of time bins examined in event  $i$ , either the bins listed in Table A.1 for the region under study, or the empty set. (For example, if the region under study is the “Lower left triangle”, then  $T(b(i))$  is the empty set for  $b(i) > 5$ .)

A chamber type  $c$  represents multiple chambers in a ring; let  $c_j$  represent the  $j$ th chamber in chamber type  $c$ , and let  $N_c$  represent the total number of chambers of type  $c$ , so that  $j = 1, \dots, N_c$ .

Let  $I_{\text{look}}(i, c_j, d)$  denote an indicator function representing whether or not we looked in wire group or half-strip  $d$  of chamber  $j$  of chamber type  $c$  in event  $i$ . That is,  $I_{\text{look}}(i, c_j, d)$  is 1 whenever  $d$  is in a half chamber in which we looked in event  $i$ , and 0 otherwise.

Let  $I_{\text{hits}}(i, c_j, d, \tau)$  denote an indicator function representing whether or not a hit is present in time bin  $\tau$  of wire group or half-strip  $d$  in chamber  $j$  of chamber type  $c$  in event  $i$ . That is,  $I_{\text{hits}}(i, c_j, d, \tau)$  is 1 whenever a hit is present, and 0 otherwise.

The product of the two indicator functions is 1 whenever we look for hits and find them. For each event  $i$ , we sum the product of the indicator functions over the time bins examined for event  $i$ , then over chambers of type  $c$ , and then further sum over events.

We thus obtain  $N_{\text{hits}}^{\text{CMS}}(c, d)$ , the number of digis found in the data set in the examined time bins of chamber type  $c$  and wire group or half-strip  $d$ :

$$N_{\text{hits}}^{\text{CMS}}(c, d) = \sum_{i=1}^{N_{\text{events}}^{\text{CMS}}} \sum_{j=1}^{N_c} \sum_{\tau \in T(b(i))} I_{\text{look}}(i, c_j, d) \times I_{\text{hits}}(i, c_j, d, \tau). \quad (\text{A.1})$$

As is evident, this sums over all chambers of type  $c$ . One may divide by  $N_c$  to get the average number of hits in one chamber of type  $c$ .

We also formalize our notation for counting the number of pp collisions to be associated with the hits found. This is complicated by the varying luminosity (or equivalently, varying pileup and/or fraction of LHC filled with bunches).

Given the association between the steady state number of pp collisions and the steady state number of hits, the number of pp collisions associated with event  $i$  is given by the mean number of pp collisions during the time interval in which we look. The time interval in which we look in one time bin  $\tau$  is the that of a time bin in the CSC readout, namely 25 ns. This is also the length in time of a bunch place in the LHC. Thus the associated number of pp collisions in event  $i$  is given by the mean number of pp collisions *per bunch place*, averaged over all bunch places at CMS (including empty bunch places), when event  $i$  was acquired. The latter is a product of two quantities: the mean number of pp collisions *per bunch crossing*  $P_i$  (also referred to as pileup) for event  $i$ , as reported by the CMS BRIL group; and the fraction of filled bunch places (bunch places that have bunch crossings), which we call  $f_{\text{fill}}$ , and which is approximately 0.62 for the data used in this analysis.

We let  $N_{\text{pp}}^{\text{CMS}}(c, d)$  be the total number of associated pp collisions in the data set at wire group or half strip  $d$  in chamber type  $c$ . It is the number of associated pp collisions, summed over all events:

$$N_{\text{pp}}^{\text{CMS}}(c, d) = \sum_{i=1}^{N_{\text{events}}^{\text{CMS}}} \sum_{j=1}^{N_c} \sum_{\tau \in T(b(i))} I_{\text{look}}(i, c_j, d) \times P_i \times f_{\text{fill}} \quad (\text{A.2})$$

As in Equation A.1, this sums over all chambers of type  $c$ , and one may divide by  $N_c$  to get the average for one chamber of type  $c$ .



The sum over  $\tau$  in Equation A.2 counts the same pp collisions more than once in an event  $i$ ; this is because we may look in more than one time bin, and the goal is to have a normalization for each time bin, and hence for hits found.

Dividing Equation A.1 by Equation A.2, we obtain the number of hits per pp collision *per wire group or half-strip*  $d$  in chamber type  $c$ :

$$\text{hits per pp per } d = \frac{N_{\text{hits}}^{\text{CMS}}(c, d)}{N_{\text{pp}}^{\text{CMS}}(c, d)}. \quad (\text{A.3})$$

To convert Equation A.3 to a per area quantity, we divide by the area subtended by each wire group or half-strip  $d$  in a chamber type  $c$ , denoted  $A_{\text{digi}}(c, d)$ . The neutron background hit rate per pp collision per area at CMS is given by

$$\text{hits per pp per area at } d = \frac{1}{A_{\text{digi}}(c, d)} \frac{N_{\text{hits}}^{\text{CMS}}(c, d)}{N_{\text{pp}}^{\text{CMS}}(c, d)} \quad (\text{A.4})$$

The average neutron background hit rate per area for the entire chamber type  $c$  is obtained by summing Equation A.3 over the set of possible wire groups and half-strips in a chamber type, denoted  $D(c)$ , and dividing by the chamber area  $A_{\text{CSC}}(c)$ :

$$\text{hits per pp per area, avg. over } c = \frac{1}{A_{\text{CSC}}(c)} \sum_{d \in D(c)} \frac{N_{\text{hits}}^{\text{CMS}}(c, d)}{N_{\text{pp}}^{\text{CMS}}(c, d)} \quad (\text{A.5})$$

To convert Equations A.4 and A.5 to per time quantities, we choose a reference luminosity  $\mathcal{L}_0$  and multiply by the mean pp collision rate per time at that luminosity,  $N_{\text{pp}}^{\text{CMS}}/t$ . That is,

$$\text{hits per pp per area per time at } d = \frac{1}{A_{\text{digi}}(c, d)} \frac{N_{\text{hits}}^{\text{CMS}}(c, d)}{N_{\text{pp}}^{\text{CMS}}(c, d)} \frac{N_{\text{pp}}^{\text{CMS}}}{t} \bigg|_{\mathcal{L}_0} \quad (\text{A.6})$$

and similarly for the number of hits in the entire chamber,

$$\text{hits per pp per area per time, avg. over } c = \frac{1}{A_{\text{CSC}}(c)} \sum_{d \in D(c)} \frac{N_{\text{hits}}^{\text{CMS}}(c, d)}{N_{\text{pp}}^{\text{CMS}}(c, d)} \frac{N_{\text{pp}}^{\text{CMS}}}{t} \bigg|_{\mathcal{L}_0} \quad (\text{A.7})$$

We obtain the mean pp collision rate,  $N_{\text{pp}}^{\text{CMS}}/t$ , via two related methods. In the first method, we multiply the measured pp interaction cross section and the instantaneous luminosity:

$$\left. \frac{N_{\text{pp}}^{\text{CMS}}}{t} \right|_{\mathcal{L}_0} = \sigma_{\text{pp}} \mathcal{L}_0 \quad (\text{A.8})$$

In the second method, we multiply the mean pileup  $P_0$  at a reference luminosity  $\mathcal{L}_0$  by the fill fraction  $f_{\text{fill}}$ , and divide by 25 ns, the bunch spacing interval.

$$\left. \frac{N_{\text{pp}}^{\text{CMS}}}{t} \right|_{\mathcal{L}_0} = \frac{f_{\text{fill}} P_0}{25 \text{ ns}} \frac{10^9 \text{ ns}}{\text{s}}. \quad (\text{A.9})$$

Both methods for measuring  $N_{\text{pp}}^{\text{CMS}}/t$  should give the same answer if we use the same cross section as used by the CMS BRIL group when computing pileup numbers [75, 76]. We favor the second method of computing  $N_{\text{pp}}^{\text{CMS}}/t$  because it uses only quantities which are directly provided by the CMS BRIL group.

In either case, the pp collision rate can in principle be scaled to other luminosities:

$$\left. \frac{N_{\text{pp}}^{\text{CMS}}}{t} \right|_{\mathcal{L}} = \left. \frac{N_{\text{pp}}^{\text{CMS}}}{t} \right|_{\mathcal{L}_0} \frac{\mathcal{L}}{\mathcal{L}_0}. \quad (\text{A.10})$$

Consistency checks are shown in Figures A.10 and A.11.

Figure A.10 is a plot of the measured  $N_{\text{pp}}^{\text{CMS}}/t$  vs. instantaneous luminosity via the second method described above. In the plot, each black dot in the plot represents a single luminosity section and the slope of this line corresponds to the pp interaction cross section,  $\sigma_{\text{pp}}$ . From Figure A.10 we read off the pp collision rate at any instantaneous luminosity. At the reference luminosity  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ,  $N_{\text{pp}}^{\text{CMS}}/t \approx 7 \times 10^8 \text{ pp/s}$ .

Figure A.11 shows a plot of instantaneous luminosity vs. mean pileup. Each dot in the plot represents a single luminosity section and is color coded according to the  $f_{\text{fill}}$  for that data taking period. Approximately 95% of the data collected corresponds to  $f_{\text{fill}} = 0.62$ . As a check that the two methods for calculating  $N_{\text{pp}}^{\text{CMS}}/t$  are equivalent, lines are drawn

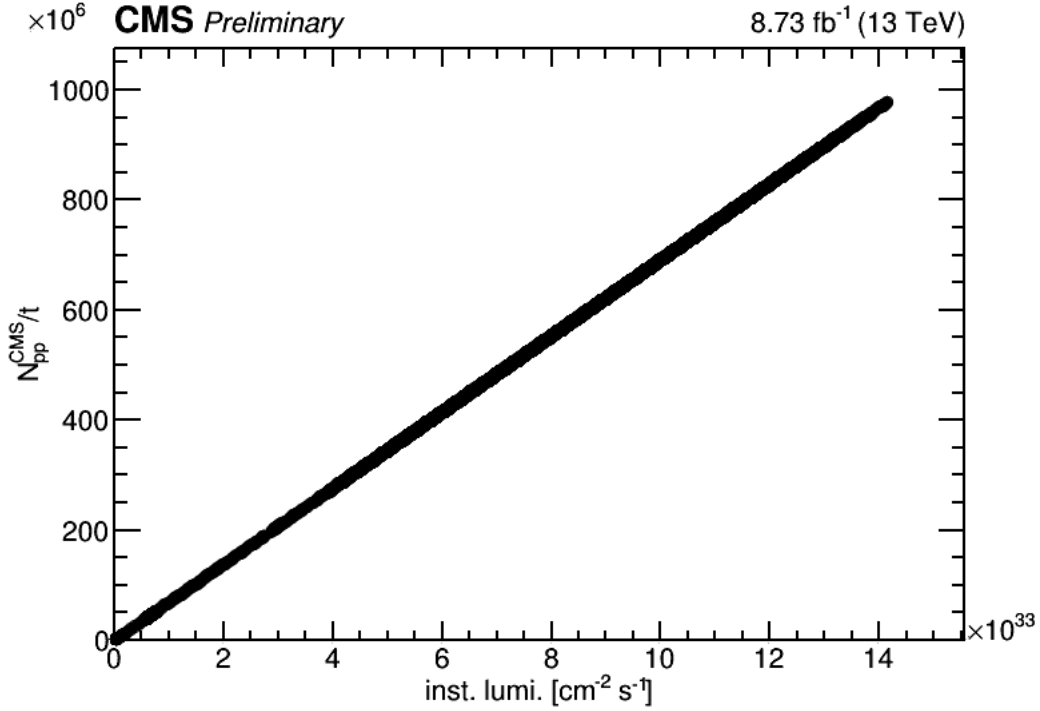


Figure A.10: Plot of  $N_{\text{pp}}^{\text{CMS}}/t$  vs. instantaneous luminosity. Each dot denotes single luminosity section.

according to the  $f_{\text{fill}}$  of various data taking conditions and the measured pp interaction cross section [77, 75].

### Computation of Neutron Hit Rate in CMS as a Function of Luminosity

The above hit rates per pp collision are independent of luminosity, and the per time rates are with respect to a reference luminosity. In this section we check the implicit assumption of linearity of rates with luminosity. We count the number of neutron hits, binned by the luminosity of the event in which the muon that generated the trigger occurred, and normalize to the number of times we look, binned by luminosity.

At the risk of confusion, we let  $\mathcal{L}$  refer to a luminosity *bin* centered on  $\mathcal{L}$ , and let  $I_{\text{lumi}}(i, \mathcal{L})$  denote an indicator function representing whether or not event  $i$  occurred with a luminosity within luminosity bin  $\mathcal{L}$ . Then the computation of hit rates within luminosity

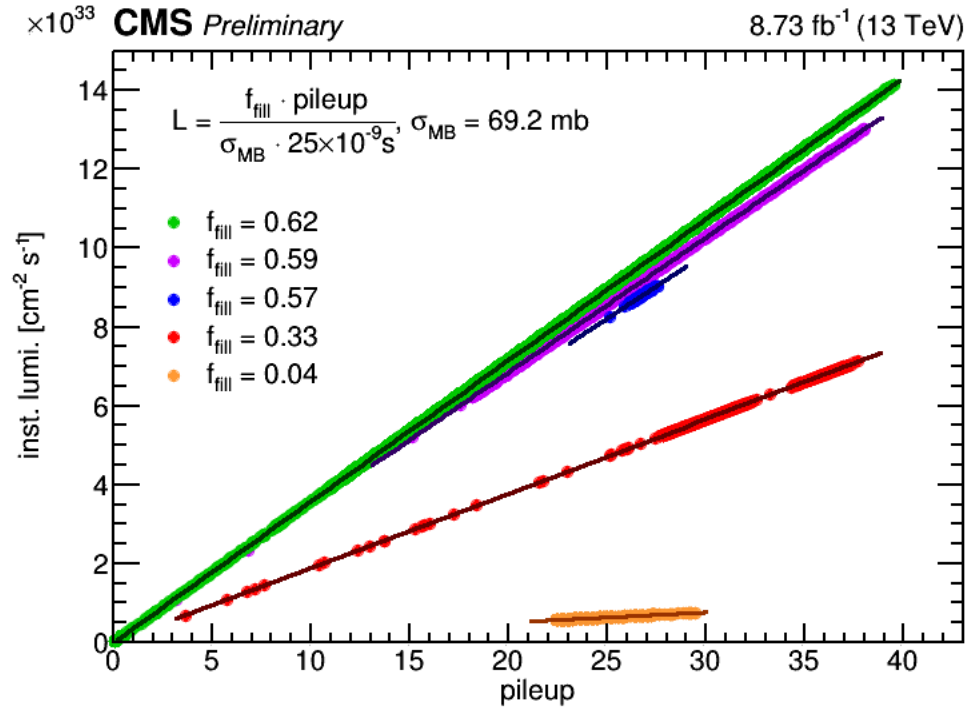


Figure A.11: Plot of instantaneous luminosity vs. mean pileup. Each dot denotes single luminosity section and the color denotes the fraction of filled bunches,  $f_{\text{fill}}$ , during which that data were collected.

bins is the same as in Equation A.1, with the addition of an indicator function to specify the luminosity bin.

We thus obtain  $N_{\text{hits}}^{\text{CMS}}(c, d, \mathcal{L})$  the number of digis found in the data set in the examined time bins of chamber type  $c$  and wire group or half-strip  $d$ , in luminosity bin  $\mathcal{L}$ :

$$N_{\text{hits}}^{\text{CMS}}(c, d, \mathcal{L}) = \sum_{i=1}^{N_{\text{events}}^{\text{CMS}}} \sum_{j=1}^{N_c} \sum_{\tau \in T(b(i))} I_{\text{look}}(i, c_j, d) \times I_{\text{hits}}(i, c_j, d, \tau) \times I_{\text{lumi}}(i, \mathcal{L}). \quad (\text{A.11})$$

The normalization for each luminosity bin,  $N_{\text{looks}}^{\text{CMS}}(c, d, \mathcal{L})$ , is the sum of the number of times we looked in the luminosity bin as in Equation A.2, without the factors counting pp collisions:

$$N_{\text{looks}}^{\text{CMS}}(c, d, \mathcal{L}) = \sum_{i=1}^{N_{\text{events}}^{\text{CMS}}} \sum_{j=1}^{N_c} \sum_{\tau \in T(b(i))} I_{\text{look}}(i, c_j, d) \times I_{\text{lumi}}(i, \mathcal{L}). \quad (\text{A.12})$$

Then we can define rates similar to those defined in Equations A.3–A.5 that are functions of luminosity:

$$\text{normalized hits per } d = \frac{N_{\text{hits}}^{\text{CMS}}}{N_{\text{looks}}^{\text{CMS}}}(c, d, \mathcal{L}) \quad (\text{A.13})$$

$$\text{normalized hits per area at } d = \frac{1}{A_{\text{digi}}(c, d)} \frac{N_{\text{hits}}^{\text{CMS}}}{N_{\text{looks}}^{\text{CMS}}}(c, d, \mathcal{L}) \quad (\text{A.14})$$

$$\text{normalized hits per area, avg. over } c = \frac{1}{A_{\text{CSC}}(c)} \sum_{d \in D(c)} \frac{N_{\text{hits}}^{\text{CMS}}}{N_{\text{looks}}^{\text{CMS}}}(c, d, \mathcal{L}) \quad (\text{A.15})$$

To convert Equations A.14 and A.15 to per time quantities, we observe that each time we look corresponds to a 25 ns time interval readout time bin, and also one bunch space. Therefore, the neutron hit rate in chamber type  $c$  as a function of luminosity is simply Equation A.15 multiplied by the conversion from time bin to ns:

$$\text{normalized hits per area per time, avg. over } c = \frac{1}{A_{\text{CSC}}(c)} \sum_{d \in D(c)} \frac{N_{\text{hits}}^{\text{CMS}}}{N_{\text{looks}}^{\text{CMS}}}(c, d, \mathcal{L}) \frac{\text{time bin}}{25 \text{ ns}} \frac{10^9 \text{ ns}}{\text{s}} \quad (\text{A.16})$$

## Computation of Neutron Hit Rate in MC Simulation

In MC simulation, we identify the neutron capture induced hits in each simulated event by the long time that is recorded since the pp collision of the event. In the rest of this section, when we refer to hits or digis, we refer to those with late enough times to be identified as being induced by neutron capture.

There is no need for a selection process as described for the CMS data (with LCTs in a corner, etc). All digis in all chambers in all simulated events are examined. We again invoke the approximation described in Section A.5 that we can associate the aggregate number of all neutron-induced hits to the aggregate number of all pp collisions, even though the pp collisions are not uniformly spaced in time at CMS. Thus we need not be concerned with recording the hit-by-hit time bin of simulated hits, but rather work with totals and averages.

At CMS, this steady-state number of neutron-induced hits per pp collision at CMS, measured in one time bin (25 ns), is the sum of hits induced by neutrons originating from all previous pp collisions. A key point is that it is not necessary to attempt such a sum, which would be awkward computationally. Rather, we note that this sum is equal to the sum of all simulated neutron-induced hits (in all future time bins) coming from a single simulated pp collision.

Let  $N_{\text{events}}^{\text{MC}}$  be the number of simulated events considered.

Let  $N_{\text{hits}}^{\text{MC}}(c, d)$  be the total number of hits in  $N_{\text{events}}^{\text{MC}}$  at wire group or half-strip  $d$  (for digi) in chamber type  $c$ . In analogy with Equation A.1, we can write, using all hits in all times,

$$N_{\text{hits}}^{\text{MC}}(c, d) = \sum_{i=1}^{N_{\text{events}}^{\text{MC}}} \sum_{j=1}^{N_c} \sum_{\text{all } \tau} 1 \quad (\text{A.17})$$

As with data, this sums over all chambers of type  $c$ . One may divide by  $N_c$  to get the average number of hits in one chamber of type  $c$ .

We let  $N_{\text{pp}}^{\text{MC}}(c)$  be the normalization to pp collisions for this hit total. (We suppress the argument  $d$  since unlike data, all wire groups and half strips are examined in every event.)

While one MC event corresponds to a single simulated pp collision, the sum over  $N_c$  chambers in Equation A.17 means that a factor of  $N_c$  is needed to consistently normalize, i.e.,

$$N_{\text{pp}}^{\text{MC}}(c) = N_{\text{events}}^{\text{MC}} N_c \quad (\text{A.18})$$

Then the number of hits per pp collision at  $d$  is then

$$\text{hits per pp per } d, \text{ MC} = \frac{N_{\text{hits}}^{\text{MC}}(c, d)}{N_{\text{pp}}^{\text{MC}}(c)} \quad (\text{A.19})$$

As in data, we convert  $N_{\text{hits}}^{\text{MC}}(c, d)/N_{\text{pp}}^{\text{MC}}(c)$  to a per area quantity by dividing by  $A_{\text{digi}}(c, d)$ . The neutron hit rate per pp collision in MC simulation is thus given by:

$$\text{hits per pp per area at } d, \text{ MC} = \frac{1}{A_{\text{digi}}(c, d)} \frac{N_{\text{hits}}^{\text{MC}}(c, d)}{N_{\text{pp}}^{\text{MC}}(c)} \quad (\text{A.20})$$

and similarly for the number of simulated hits in the entire chamber,

$$\text{hits per pp per area, avg. over } c, \text{ MC} = \frac{1}{A_{\text{CSC}}(c)} \frac{1}{N_{\text{pp}}^{\text{MC}}(c)} \sum_{d \in D(c)} N_{\text{hits}}^{\text{MC}}(c, d) \quad (\text{A.21})$$

Finally, we convert Equation A.20 and A.21 to per time quantities by choosing a reference luminosity  $\mathcal{L}_0$  and multiplying by the pp collision rate per time at that luminosity,  $N_{\text{pp}}^{\text{MC}}/t$ :

$$\text{hits per pp per area per time at } d, \text{ MC} = \frac{1}{A_{\text{digi}}(c, d)} \frac{N_{\text{hits}}^{\text{MC}}(c, d)}{N_{\text{pp}}^{\text{MC}}(c)} \frac{N_{\text{pp}}^{\text{MC}}}{t} \quad (\text{A.22})$$

and

$$\text{hits per pp per area per time, avg. over } c, \text{ MC} = \frac{1}{A_{\text{CSC}}(c)} \frac{1}{N_{\text{pp}}^{\text{MC}}(c)} \sum_{d \in D(c)} N_{\text{hits}}^{\text{MC}}(c, d) \frac{N_{\text{pp}}^{\text{MC}}}{t} \quad (\text{A.23})$$

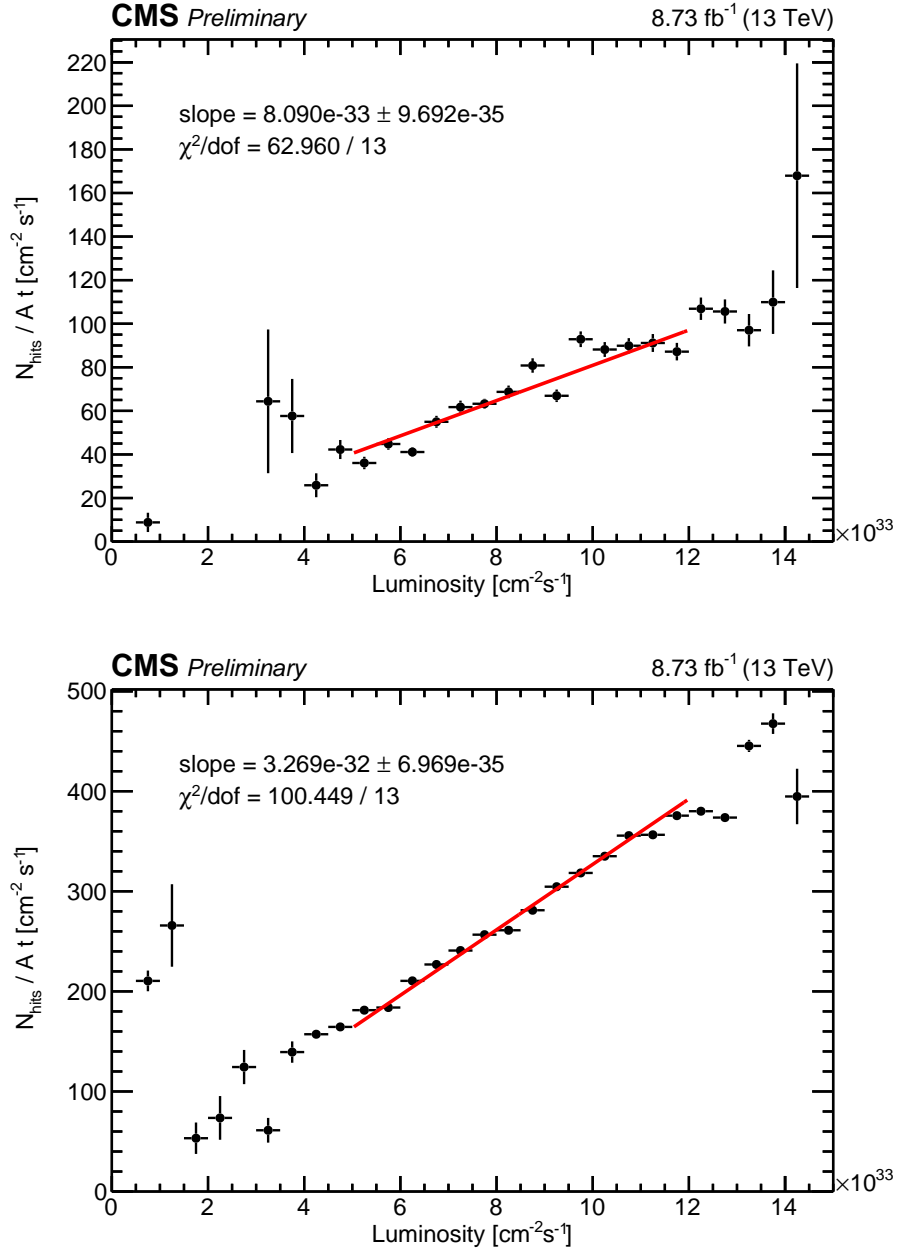


Figure A.12: Plot of hits per time per area as a function of luminosity as calculated in Equation A.16 for ME1/1 chambers, (top) for hits that occur at the end of LHC gaps (candidate neutron capture induced hits) and (bottom) for hits that occur during pp collisions. The red line is a linear fit constrained to go through the origin and fit over the central region of luminosity.



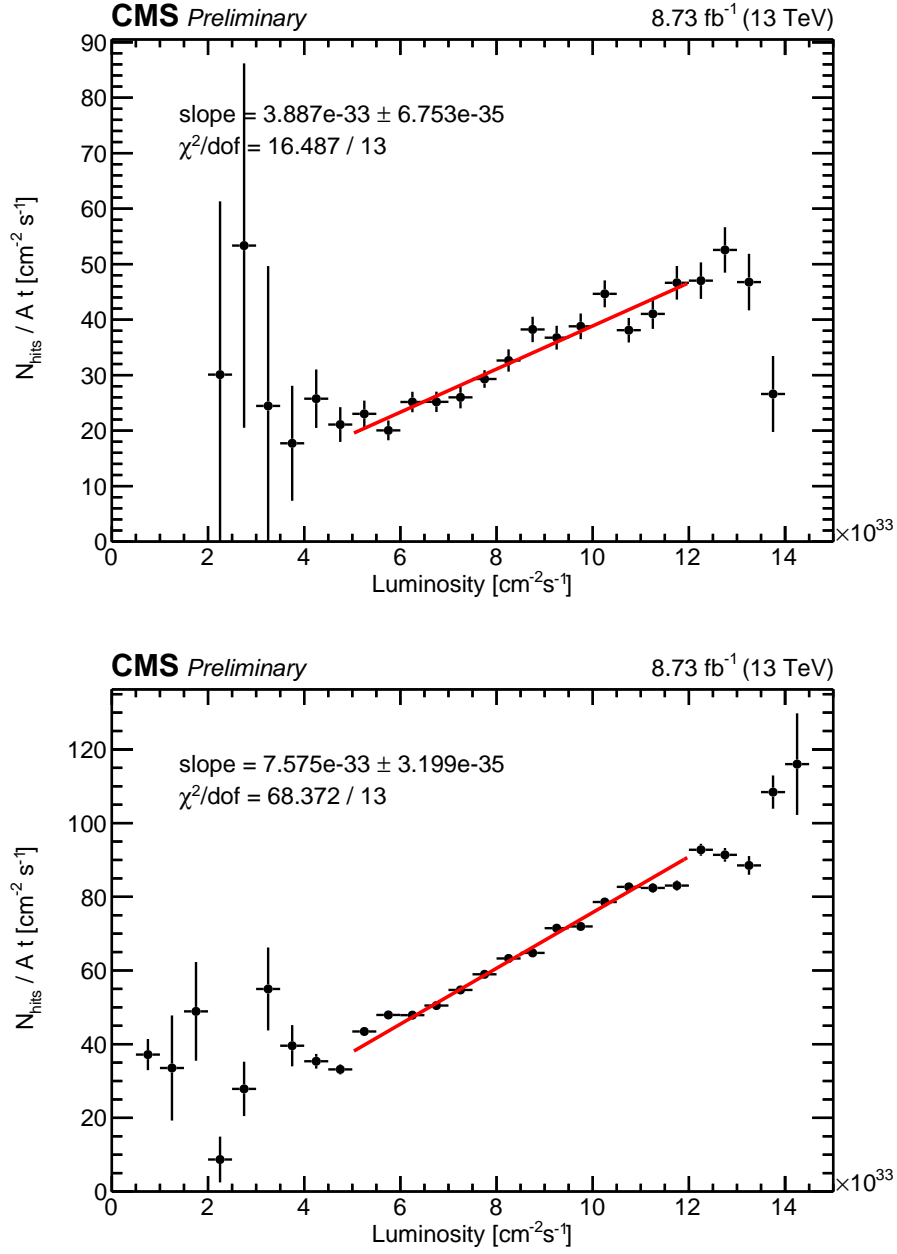


Figure A.13: Plot of hits per time per area as a function of luminosity as calculated in Equation A.16 for ME2/1 chambers, (top) for hits that occur at the end of LHC gaps (candidate neutron capture induced hits) and (bottom) for hits that occur during pp collisions. The red line is a linear fit constrained to go through the origin and fit over the central region of luminosity.

## Results

Figures A.12 and A.13 display plots of neutron-induced wire group hit rate without normalizing to pp collisions (as in Equation A.16) vs. luminosity, for hits that occur at the end of LHC bunch gaps (lower left triangle from Figure A.9) and hits that occur during CMS pp collisions (rectangle in Figure A.9), for ME1/1 and ME2/1 chambers, respectively. We plot the hit rate that is not normalized to pp collisions, a quantity that is a function of luminosity, in order to see any potential linear dependence of hits vs. luminosity. Indeed, these plots show a relationship that is linear to the eye and intercepts the origin. This suggests that we have succeeded in isolating neutron capture induced hits, because contamination from muon-induced background would result in a positive offset on the  $y$  axis. Fluctuations in the data points are presumably due to changes in data taking conditions; however, the size of the  $\chi^2/\text{dof}$  from the fit suggest that there are some systematic uncertainties that are not understood.

Figures A.14 and A.15 display histograms of the neutron-induced wire group hit rate for CMS data and MC, Equations A.6 and A.22, as a function of wire group number, in ME1/1 and ME2/1 chambers, respectively. We plot the hit rates that have been multiplied by  $N_{\text{pp}}^{\text{MC}}/t$  for a reference luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  so as to compare data and MC, as in Equations A.6 and A.22. Rates normalized to area and time for both MC simulation and CMS data are shown, for both the XS and HP cross section libraries. The agreement is good to a factor of 2, depending on chamber type. The agreement in the HP plot of ME2/1 is anomalously good, presumably by chance.

In preliminary results in 2017 for the neutron induced hit rates, we incorrectly normalized the number of hits we counted with respect to the number of pp collisions in CMS data. Specifically in Equation A.2, we omitted the fill factor, which changes the result for data by about a factor of 0.62 with respect to MC simulation. In addition, the way we expressed the rates per time was ambiguous and confusing; we trust that the current methodology is more clear.

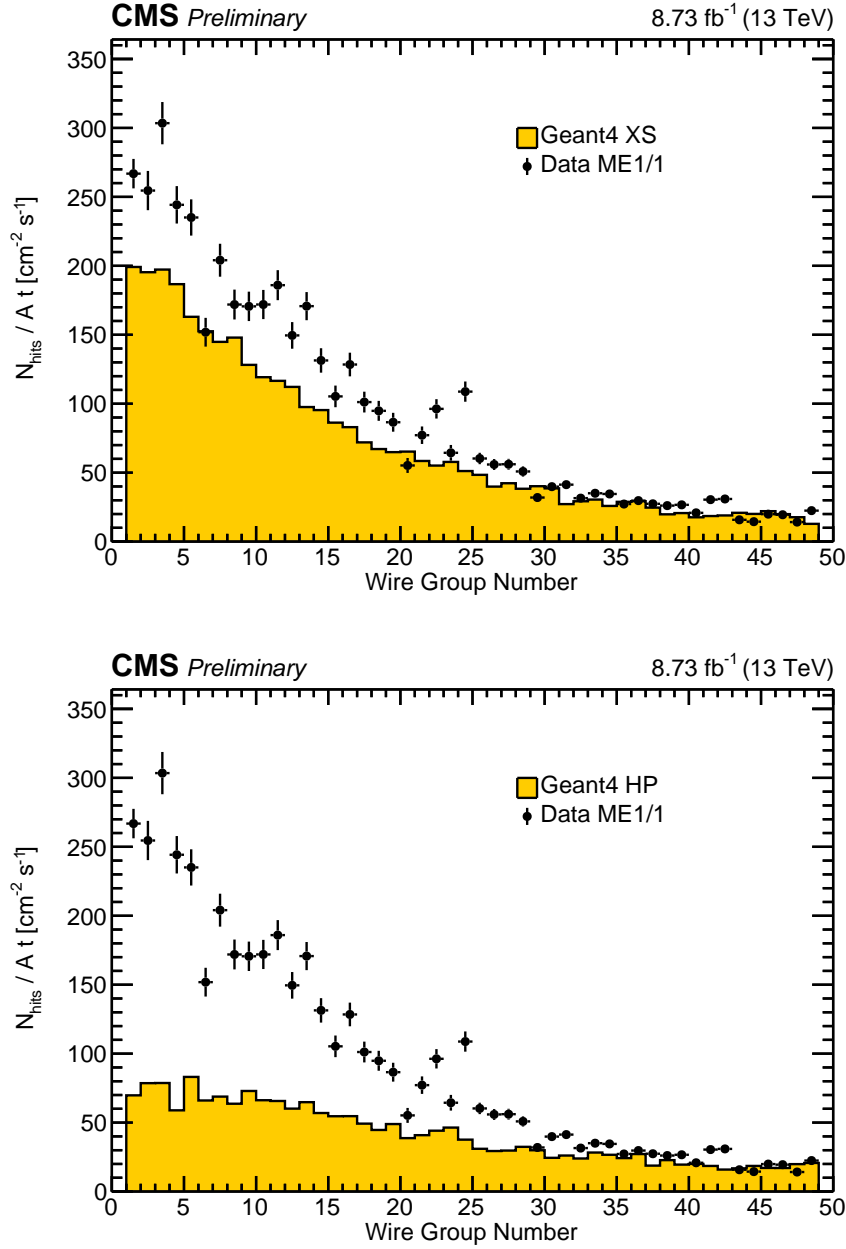


Figure A.14: Histogram of neutron capture induced anode wire hits per time per area for CMS data (as calculated in Equation A.6) and for MC simulation (as calculated in Equation A.22) for ME1/1 chambers for a reference luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . CMS data are compared to results from GEANT4 (top) XS and (bottom) HP neutron interaction cross section libraries in CMS MC simulation of minimum-bias proton-proton collisions at 13 TeV.

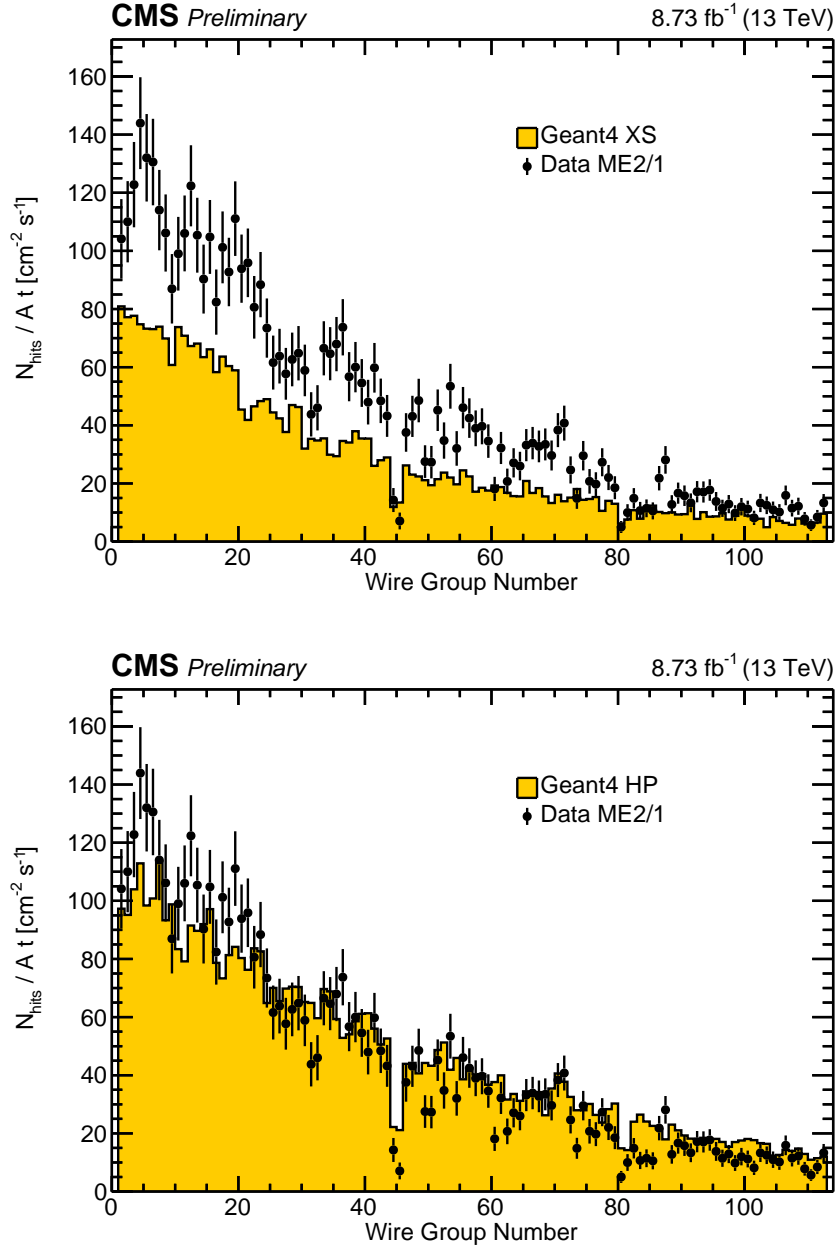


Figure A.15: Histogram of neutron capture induced anode wire hits per time per area for CMS data (as calculated in Equation A.6) and for MC simulation (as calculated in Equation A.22) for ME2/1 chambers for a reference luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . CMS data are compared to results from GEANT4 (top) XS and (bottom) HP neutron interaction cross section libraries in CMS MC simulation of minimum-bias proton-proton collisions at 13 TeV.

## Neutron Hit Patterns

A neutron-induced electron that leaves hits in more than one strip or layer is potentially more disruptive to reconstruction than one that leaves a single isolated hit. It is of interest to see if the MC simulation reproduces digi patterns of single, double, and triple hits found in CMS data. From the selected neutron digis, we consider contiguous clusters of at most  $3 \text{ half-strips} \times 3 \text{ layers}$ , i.e.  $\phi$ - $z$  patterns, and study patterns for both CMS data and MC simulation.

Figure A.16 displays histograms of the neutron-induced hit  $\phi$ - $z$  pattern distribution for CMS data and MC simulation. The  $x$  axis bin labels each show a representation of the half-strip vs. layer pattern; the horizontal direction in each is half-strips, and the vertical direction in each is layers. The sections of plots are colored by the number of hits in each pattern; on the far left in green is the only 1-hit pattern, followed by all possible 2-hit patterns in blue, and finally all possible 3-hit patterns in orange. The shaded histogram bins indicate patterns that have comparators on adjacent half-strips on the same layer. Since trigger electronics only report comparator hits on adjacent half-strips if they are in different time bins, the occupancy of these patterns should be suppressed compared to patterns to which they are otherwise geometrically similar. Comparing the shaded histogram bins in Figure A.16 to the unshaded histogram bins suggests that this is indeed the case.

The distribution of neutron-induced hit patterns is overwhelmingly single, isolated hits. Both the occupancy of each pattern type and the relative ratio between singles, doubles, and triples, as well as to each other, show rough agreement between CMS data and MC simulation.

## A.6 Muon Test Beam Studies at GIF++

### The GIF++ Experimental Setup

The CERN Gamma Irradiation Facility (GIF++) is located on the CERN Prévessin site in the H4 beam line extracted from the Super Proton Synchrotron (SPS) [78]. The SPS

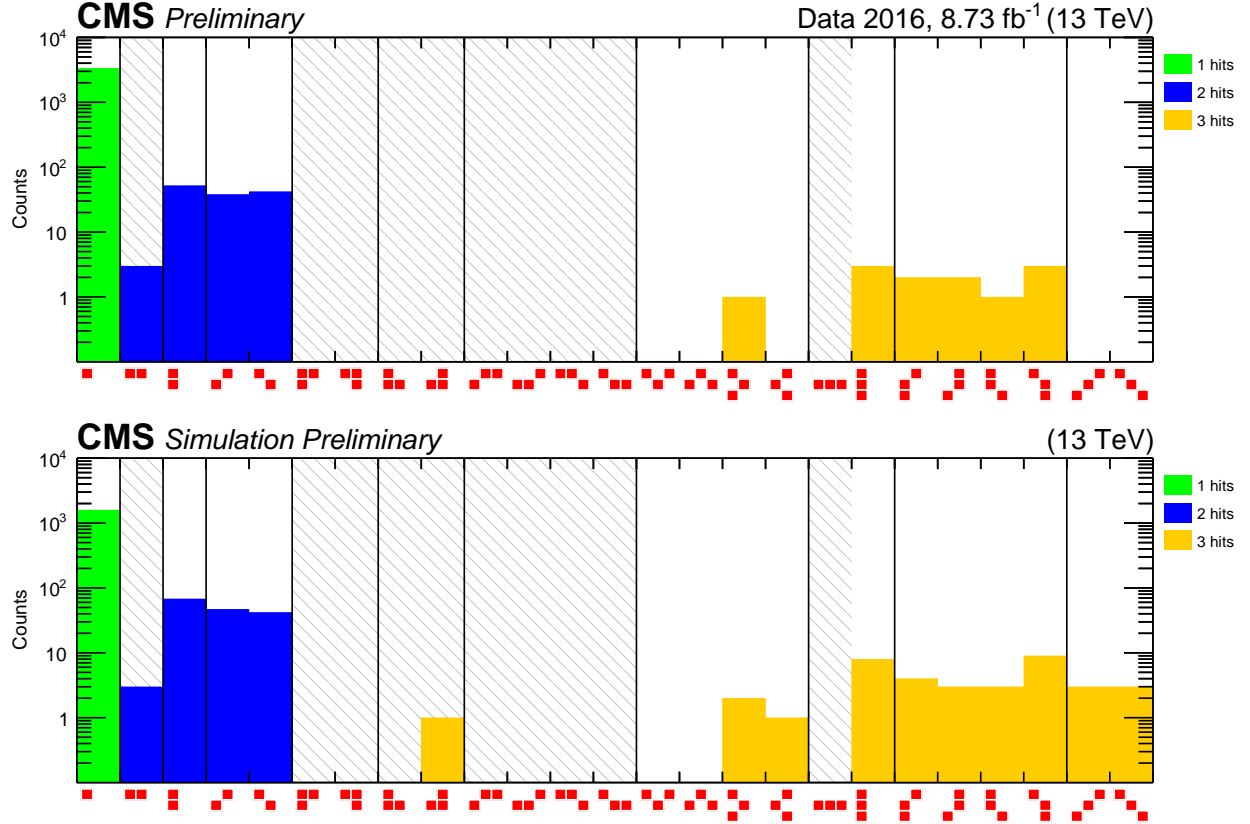


Figure A.16: Distribution of candidate neutron capture induced patterns in (top) 2016 CMS data and (bottom) CMS simulation. The patterns shown here are 1, 2, and 3-hit clusters only. For each red 3×3 pattern, the horizontal direction represents cathode half-strips and the vertical direction represents a layer in the CSC. The red boxes in a pattern indicate the presence of a candidate neutron capture induced hit in a cathode half-strip and layer. The shaded histogram bins correspond to patterns that are suppressed by CSC firmware electronics.

delivers a beam of 400 GeV protons on a fixed target, providing a charged particle beam of hadrons decaying to a beam primarily consisting of muons with a broad momentum spectrum around 100 GeV. The facility includes a 13.9 TBq <sup>137</sup>Cs gamma ray irradiation source emitting primarily 662 keV photons with the means to attenuate the source over a large range. Two CSCs (one ME1/1 and one ME2/1) are placed within the facility. This setup is used for studying aging, gas gain, high-luminosity conditions, etc., in preparation for the HL-LHC.

Level-1 Trigger Accepts (L1As) from muons, triggering chamber readout, can be produced in one of two ways. The first way is by self-trigger, where L1As are produced by chamber electronics forming Local Charged Tracks (LCTs) using coarsely correlated anode wire and cathode strip hits. Variables under control in the experimental setup include the chamber anode high voltage, the gamma ray source intensity (via attenuating filters, allowing a range of intensities from fully open to nearly closed as well as completely off), and the firmware of the controlling electronics, which defines various settings such as hit multiplicity thresholds. The second way to produce L1As is by external trigger, where L1As are produced by the triple coincidence of three scintillators in the beam: one upstream of the source, one downstream of the source, and one that is mounted in the path of the beam just in front of the chambers. The GIF++ data used in this note are externally triggered data taken during a muon test beam over a large range of GIF++ source intensities with CMS-like firmware parameters, and with a set of high voltages intended to equalize gas gain across all layers.

### **LCT Efficiency and RecHit Displacement at GIF++**

To measure the LCT efficiency, we begin by counting the fraction  $f$  of L1As that have at least one LCT within the area shadowed by the middle scintillator mounted in front of the chambers. Due to issues of precisely defining the shadow and dealing with events with more than one muon, we do not attempt to define an absolute efficiency. Instead, we define a relative efficiency dividing the value of  $f$  taken at higher source intensities by the value of  $f$  with the GIF++ source off, denoted  $f_{\text{off}}$ . We estimate the scintillator area boundary empirically using a position scatter plot of LCTs, and count those LCTs whose layer 3 half-strip and wire group number fall within this boundary. Figure A.17 displays plots of the LCT relative efficiency,  $f/f_{\text{off}}$ , as a function of chamber anode HV current induced by increasing the GIF++ source intensity for the ME1/1 and ME2/1 chambers at GIF++. As the gamma source intensity is increased, the number of photon-induced noise hits increases, and the number of LCTs constructed decreases.

One possible cause for the loss of signal muon LCTs is the corruption of muon hits

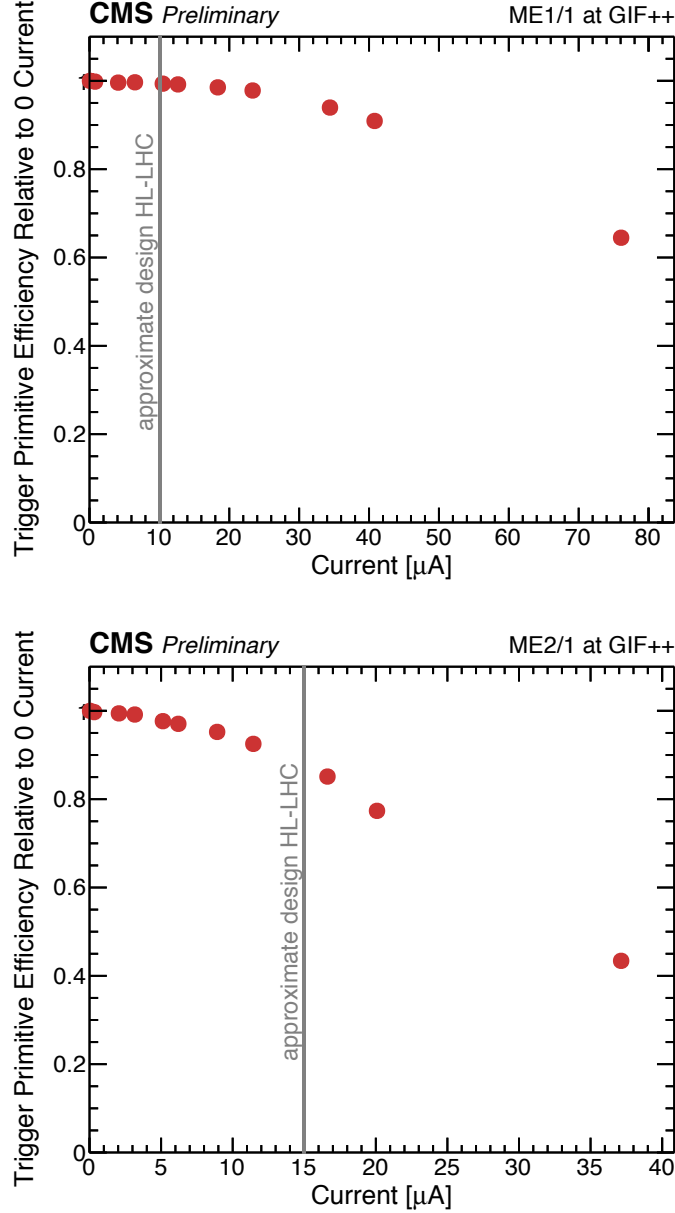


Figure A.17: Plot of LCT efficiency divided by the efficiency at minimum HV current (irradiation source turned off), for the (top) ME1/1 and (bottom) ME2/1 chambers at GIF++. LCT efficiency is defined as the number of LCTs created within a scintillator shadow, divided by the number of scintillator triggers received. As the source intensity is increased (chamber anode HV current is increased), the LCT efficiency drops. The gray line indicates an approximate equivalent HV current corresponding to design HL-LHC luminosity.



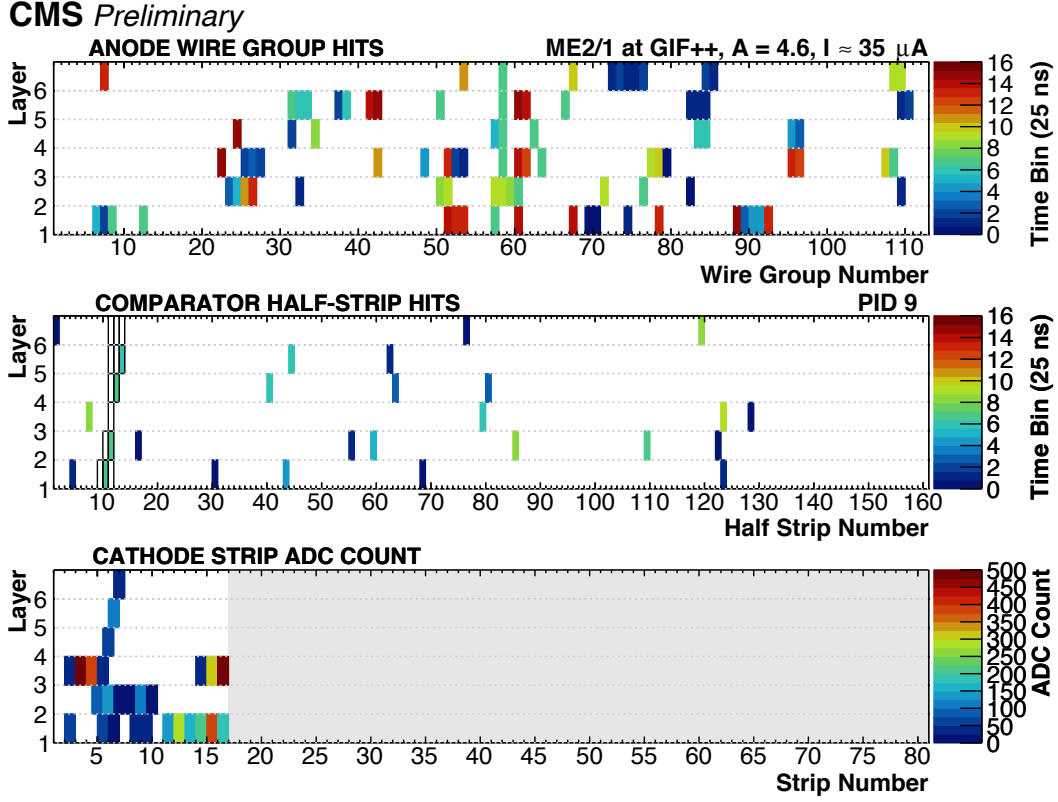


Figure A.18: Display of an event collected during a muon test beam at the CERN Gamma Irradiation Facility (GIF++) with a CSC from ME2/1, showing digitized detector responses: anode wire ( $r$  coordinate) responses (wire group hits), cathode half-strip ( $\phi$  coordinate) responses (comparator half-strip hits), and cathode strip analog-to-digital-converter (ADC) counts proportional to deposited charge. Each display is organized by the gas gap layer and the strip or wire number in which the response occurred. The quantity A represents the attenuating factor applied to the 13.9 TBq  $^{137}\text{Cs}$  gamma irradiation source and I is the chamber anode wire HV current. This display illustrates a mechanism by which a photon hit can displace a muon hit; the large amount of deposited charge seen at the left edge of the ADC counts resulted in a corresponding shifted comparator hit.

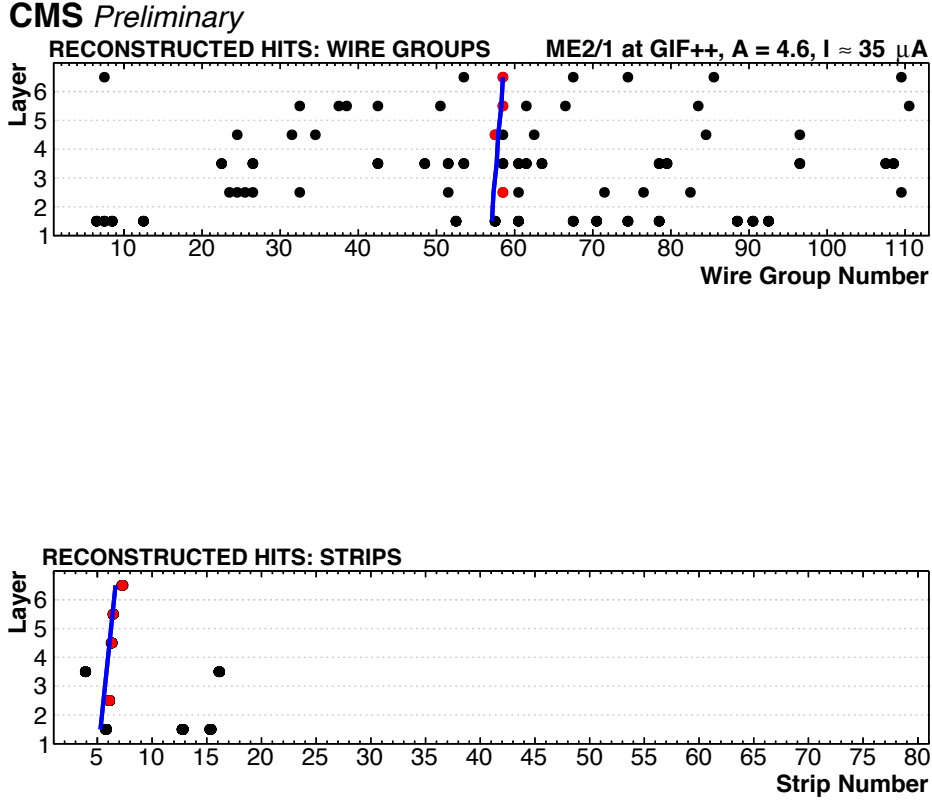


Figure A.19: Display of the same event as Figure A.18, showing offline reconstructed detector responses (RecHits), projected along the anode wire axis and the cathode strip axis (black dots). Blue lines indicate the offline reconstructed muon segment created by a straight-line fit to the red highlighted RecHits. This display illustrates a mechanism by which a photon hit can displace a muon hit; the reconstructed hit in layer 3 was displaced and subsequently excluded from the segment fit.

induced by background. To understand the characteristics of hit corruption, we examine CSC event displays such as in Figure A.18. These event displays show wire group hits, comparator half-strip hits, and strip ADC counts with the layer on the vertical axis and the digi number on the horizontal axis. The color for the wire groups and comparators indicates the digi time bin, while the color for the strip ADC counts indicates the ADC counts. The comparator hit plot has an overlay in black rectangular outline of the LCT pattern that was used to identify the LCT.

This particular event display illustrates a way that a photon-induced hit may cause loss of LCTs, and consequently, loss of offline reconstructed segments. The display shows a muon triggering readout and forming a 4-hit LCT, whose comparator pattern can be seen in half-strips 9–13. The strip ADC counts show a large energy deposit that we attribute to a photon-induced hit in strips 3–4 in layer 3. The corresponding comparator hit in half-strip 7 is shifted away from what would have been its correct position in approximately half-strip 11. This results in what would have been a 5-hit LCT deteriorating to a 4-hit LCT. This ability of photon hits to displace hits can in this way cause loss of LCTs as well as deterioration of their quality.

Figure A.19 is a display of the corresponding offline reconstructed hits (RecHits) [79], along with an overlay of the muon segment constructed from them. The RecHits that contributed to the segment are shown in red. The shifted comparator hit and the photon energy deposit resulted in a RecHit that was displaced to the left, resulting in its exclusion from the segment. This directly results in a deterioration of segment reconstruction resolution.

Figure A.20 displays stacked plots of the fraction of reconstructed muon segments spatially closest to muon LCTs within the scintillator acceptance area, categorized by the number of hits used to form that segment, as a function of chamber anode HV current for the ME1/1 and ME2/1 chambers at GIF++. The plot shows that the fraction of high quality segments formed with 6 hits decreases as the source intensity is increased, while the fraction of lower quality segments formed by 3 hits increases. This confirms that LCT loss can be caused by a loss of hits, and demonstrates the mechanism by which neutron-induced hits can potentially disrupt CSC detector performance.

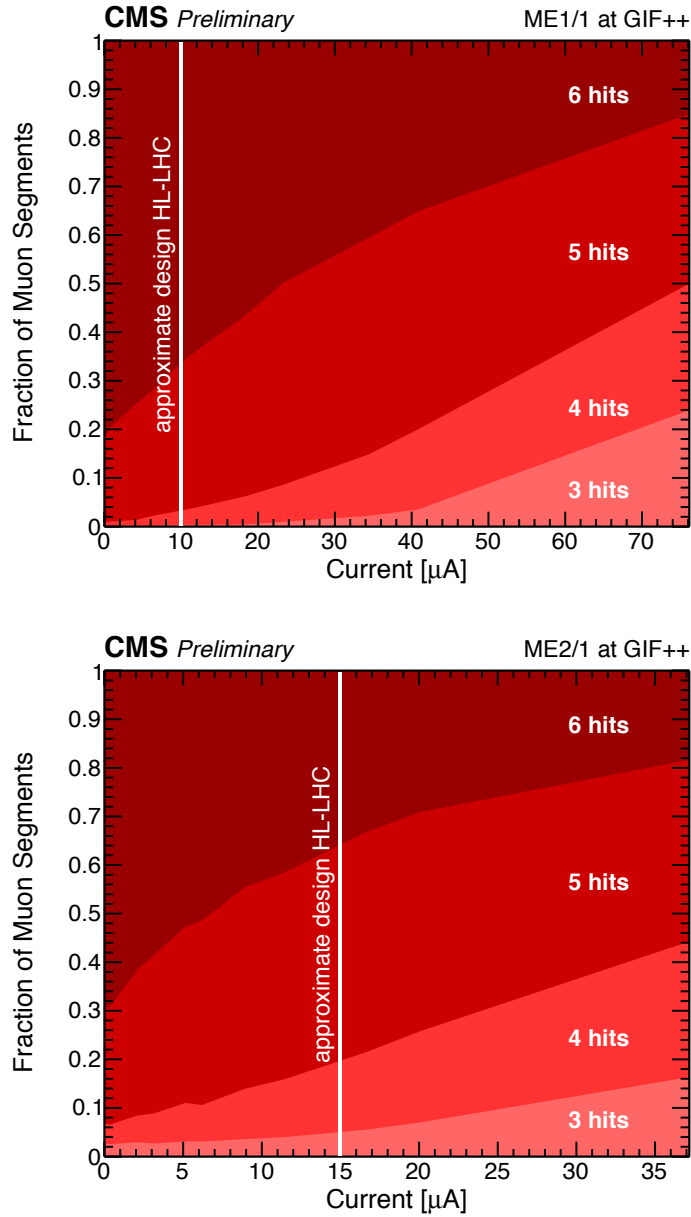


Figure A.20: Stacked plot of fraction of offline reconstructed muon segments, categorized by the number of reconstructed hits used to form them, vs. chamber anode HV current, for the (top) ME1/1 and (bottom) ME2/1 at GIF++. As the source intensity is increased (chamber anode HV current is increased), the fraction of good quality 6 hit segments decreases, while the fraction of lower quality 3 hit segments increases. The white line indicates an approximate equivalent HV current corresponding to design HL-LHC luminosity.

## A.7 Charge per Hit Studies at GIF++ and in CMS Data and Simulation

As discussed in Section A.6, the GIF++ facility allows us to study the performance of CSC online and offline reconstruction under conditions that potentially correspond to the radiation environment expected at the HL-LHC. To investigate the correspondence between GIF++ source intensity and LHC instantaneous luminosity, we examine the CSC anode HV current as measured by an ammeter in the CSC HV power supply (HVPS) of both CSCs at GIF++ and for similar chamber types at CMS.

Differences in radiation environments produced from the GIF++ source and from LHC pp collisions necessitate understanding the validity of using anode HV current for the correspondence between GIF++ and CMS. As noted in Section A.6 above, GIF++ photons are emitted with a maximum energy of 662 keV by the  $^{137}\text{Cs}$  source, and the resulting ionizing electrons are primarily from Compton scattering and the photoelectric effect. In contrast, photons originating from the neutron background in CMS can carry up to several MeV in energy and thus, in addition to more energetic electrons from Compton scattering and the photoelectric effect, ionizing electrons and positrons can arise from pair production. Because of these differences and to explore any others, we measure the charge/hit of background hits that are expected to be the dominant source of charge in the anode HV current by measuring the ratio of the anode HV current to the anode wire group hits/s in both CMS data and GIF++ data.

### Computation of charge/hit from CMS data

#### ME1/1

To calculate charge/hit in ME1/1 chambers using CMS data, we take the ratio of the measured HV current (charge/s) and the anode wire group hits/s. Each of the measured HV current and anode wire group hits/s are calculated at a fixed reference luminosity intended to be representative of a typical measured values of the HV current and hits/s. We choose

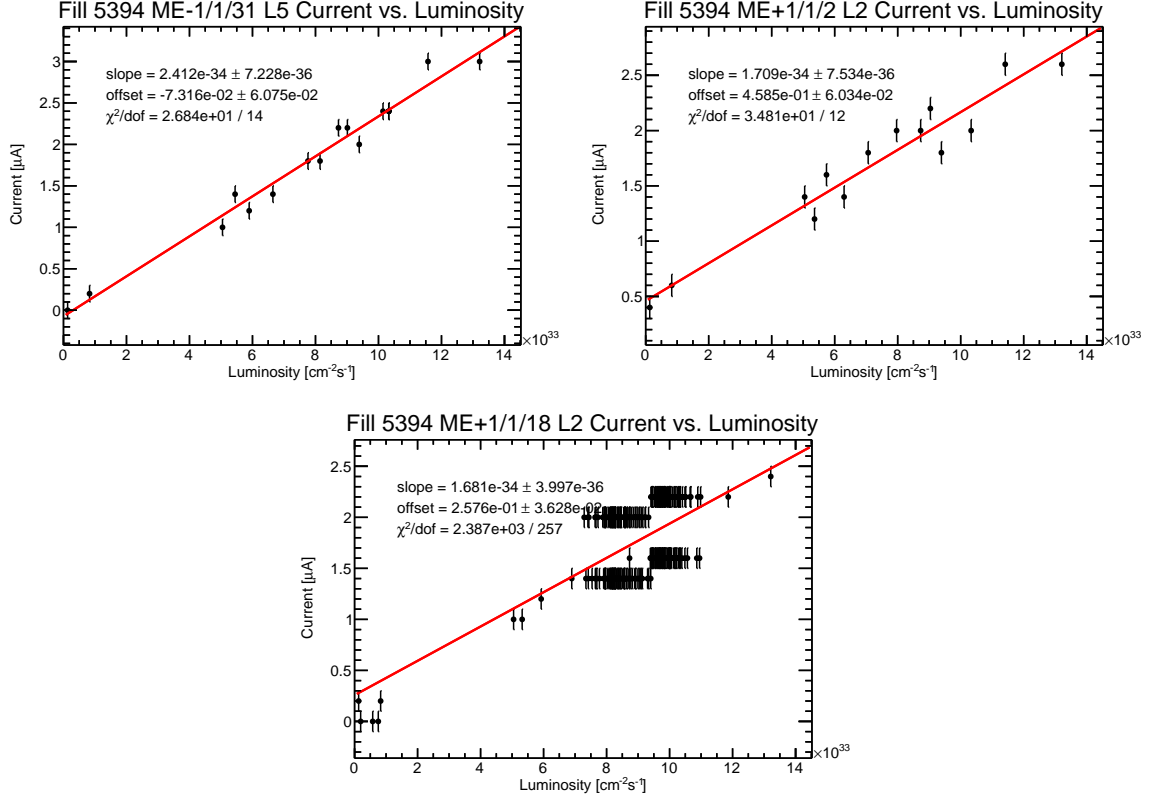


Figure A.21: Plot of anode HV current vs. LHC luminosity with a two-parameter linear fit for (top) a well-behaved ME1/1 channel, (middle) an ME1/1 channel with a large fitted offset, and (bottom) a noisy ME1/1 channel.

$10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  as the reference luminosity.

We use anode HV current data collected by the HVPS and stored in an external database during LHC fills to calculate the HV current at the reference luminosity, which we refer to as the reference HV current. We read the HV current data offline from the database and correlate each HV current measurement by time to the corresponding measured LHC luminosity. In ME1/1 chambers, each layer constitutes an HV channel. We retrieve the HV current measurements for all ME1/1 HV channels, for all LHC fills during data taking period Run 2016H. We then study plots of anode HV current vs. luminosity and attempt to calculate the reference HV current by performing a two-parameter linear fit to the data and reading off the value of the fit at  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ .

However, since the HVPSs for the ME1/1 chambers at CMS are known to measure a non-zero offset of HV current at zero luminosity, some adjustment of the HV current is necessary. In addition, in some ME1/1 HV channels, the HVPS can produce very noisy measurements of the anode HV current, resulting in data that are inconsistent with a straight line fit. Figure A.21 displays three plots of anode HV current vs. luminosity. The top plot is an example of a well-behaved ME1/1 channel that does not have an offset in the vertical intercept, so that we can use it to calculate the reference HV current. The middle plot is an example of ME1/1 channels that have a large offset of  $0.46\,\mu\text{A}$  that needs to be corrected when calculating the reference HV current. The bottom plot is an example of a noisy HV channel where the HVPS reads out hundreds of values of the HV current within a short period of time where the measured values oscillate between two values separated by  $0.6\,\mu\text{A}$  and is not consistent with the model of a straight line fit.

To avoid problems such as large offsets and many oscillating values, we scanned through all 432 ME1/1 HV channels in LHC fill 5394. We selected what we thought were the best channels, similar in appearance to the top plot in Figure A.21, based on fit  $\chi^2$ , number of degrees of freedom, and our own judgment. We applied this list of 157 (out of 432 possible) selected channels for use in all other LHC fills considered. We considered a total of 23 fills in this study.

Figure A.22 is a histogram of reference HV current values at the reference luminosity from hand-selected channels with the offset required to be less than  $\pm 0.1\,\mu\text{A}$ . To obtain one characteristic HV current value for each fill, we take the average of the plus and minus endcaps. Then to combine the measurements over all fills, we take the average of the HV currents in each fill during Run 2016H, with each fill weighted by its integrated luminosity. The averaged HV current is then multiplied by 6 because each CSC chamber has six HV channels (one per chamber layer) which operate independently. The resulting total ME1/1 chamber HV current is

$$I_{\text{ME1/1}} = 11.3\,\mu\text{A} \tag{A.24}$$

at the reference luminosity.

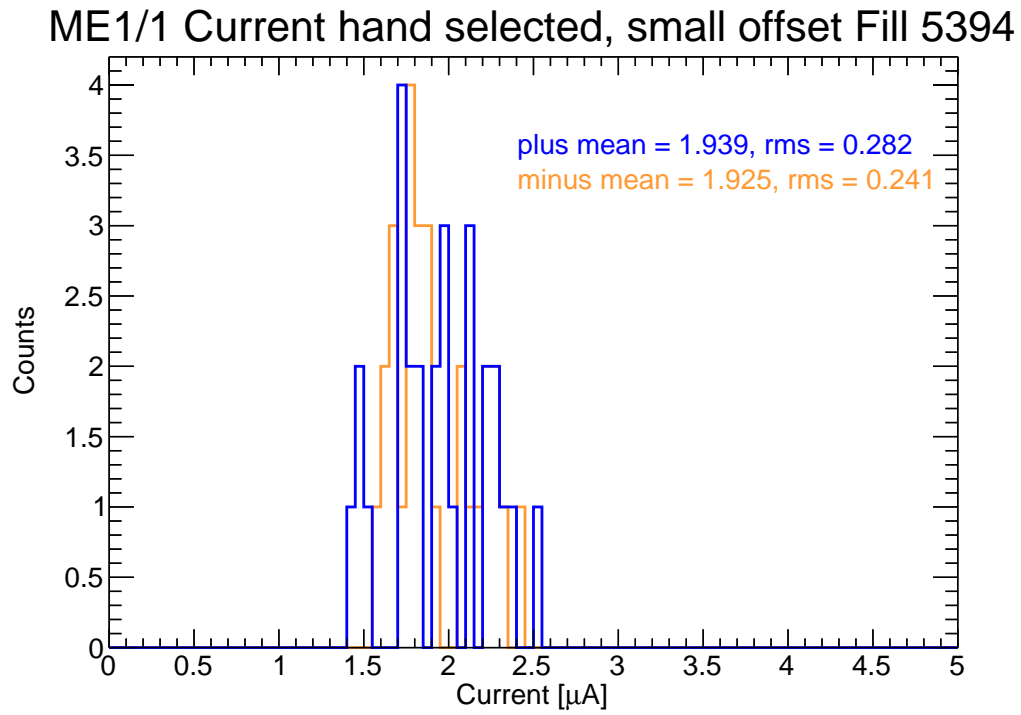


Figure A.22: Histogram of offset-corrected reference HV currents from ME1/1 HV channels that pass the hand selection color-coded by endcap.



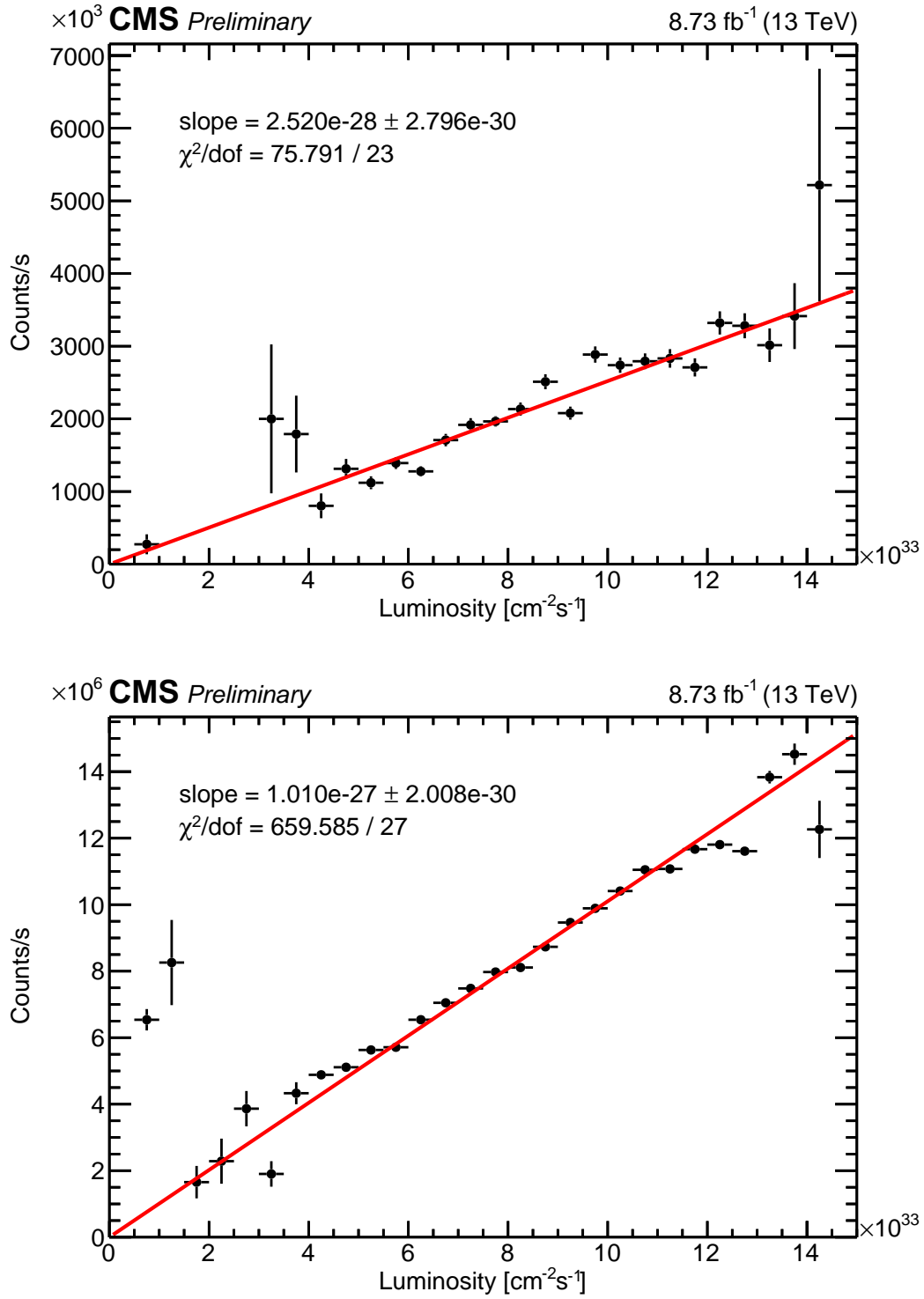


Figure A.23: Plot of ME1/1 wire group  $N_{\text{hits}}^{\text{CMS}}/t$  vs. LHC luminosity with a one-parameter linear fit for hits that occur (top) during LHC bunch gaps and (bottom) during pp collisions.

To obtain the reference rate of anode wire group hits, we start by calculating the values of the hits/s at the reference luminosity  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  from the slopes of one-parameter linear fit to plots of hits/s vs. luminosity in Figure A.23. The top plot in Figure A.23 is the ME1/1 wire group hits/s vs. LHC luminosity (from Equation A.7 but without the  $1/A_{\text{CSC}}$  factor) for wire group hits that occur in LHC bunch gaps (lower left triangle from Figure A.9); the bottom plot is the same but for wire group hits that occur during pp collisions (middle rectangle from Figure A.9). The reference rate of anode wire group hits/s is sum of the hit rates of hits occurring in LHC gaps,  $H_{\text{LHC gap}}$ , and during LHC bunch trains,  $H_{\text{pp-coll}}$ , weighted by the LHC bunch fill fraction,  $f_{\text{fill}} = 0.62$ ,:

$$\begin{aligned}
H_{\text{WG-ME1/1}} &= f_{\text{fill}} H_{\text{pp-coll}} + (1 - f_{\text{fill}}) H_{\text{LHC gap}} \\
&= 0.62 \cdot 10.1 \times 10^6 \text{ hits/s} + 0.32 \cdot 2.5 \times 10^5 \text{ hits/s} \\
&= 7.2 \times 10^6 \text{ hits/s}.
\end{aligned} \tag{A.25}$$

The charge/hit is then calculated by dividing the anode HV current by the anode wire group hits/s:

$$\text{charge/hit} = I_{\text{ME1/1}}/H_{\text{WG-ME1/1}} = (11.3 \mu\text{A})/(7.2 \times 10^6 \text{ hits/s}) = 1570 \text{ fC/hit}. \tag{A.26}$$

## ME2/1

ME2/1 chambers each have 18 independent HV channels. Each of a chamber's six layers is divided into three HV segments: one HV segment, through which passes the highest current, is close to the beam line; one HV segment, through which passes a lower current, is in the center of the chamber; and one HV segment, through which passes the lowest current, is at the end of the chamber furthest from the beam line. Figure A.24 displays plots of the anode HV current vs. luminosity for a representative example of ME2/1 chambers, ME+2/1/04 Layer 5, for all three HV segments.

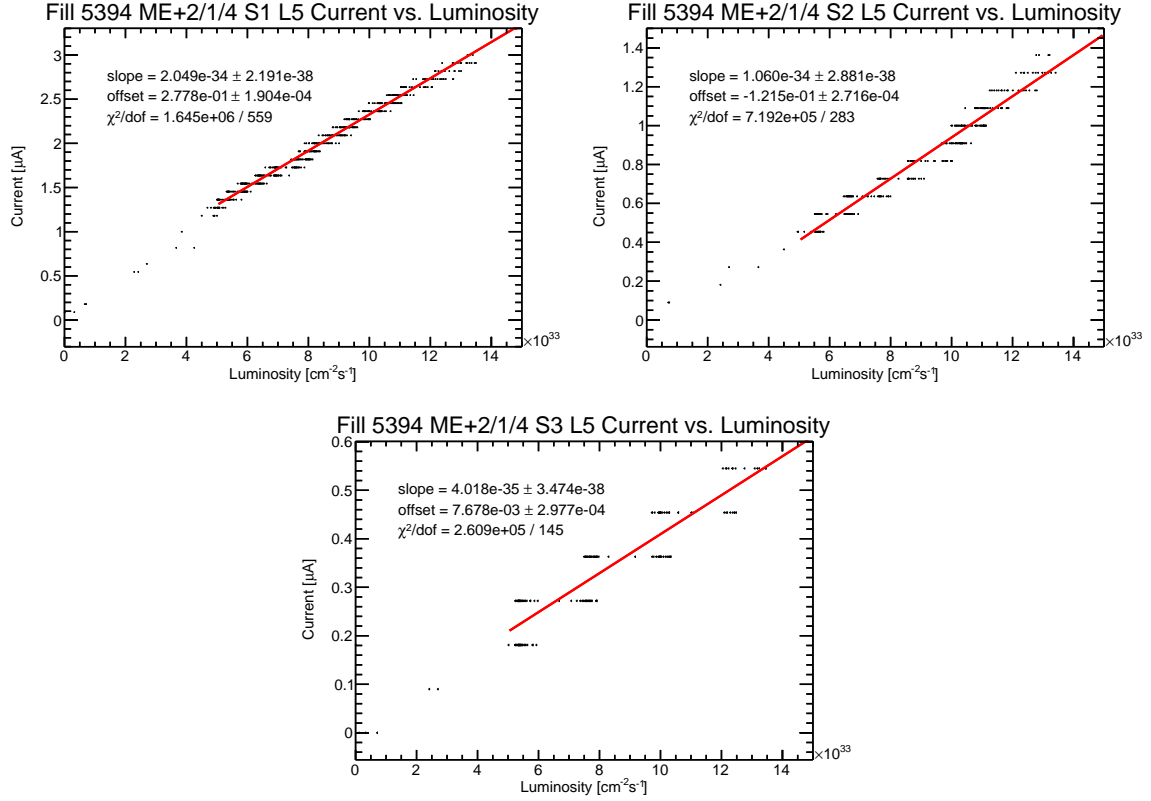


Figure A.24: Plot of current vs. luminosity for ME+2/1/04 layer 5 for HV Segment (top) 1, (middle) 2, and (bottom) 3, with a two-parameter linear fit. The least count of approximately  $0.1 \mu\text{A}$  is most evident in HV segment 3.

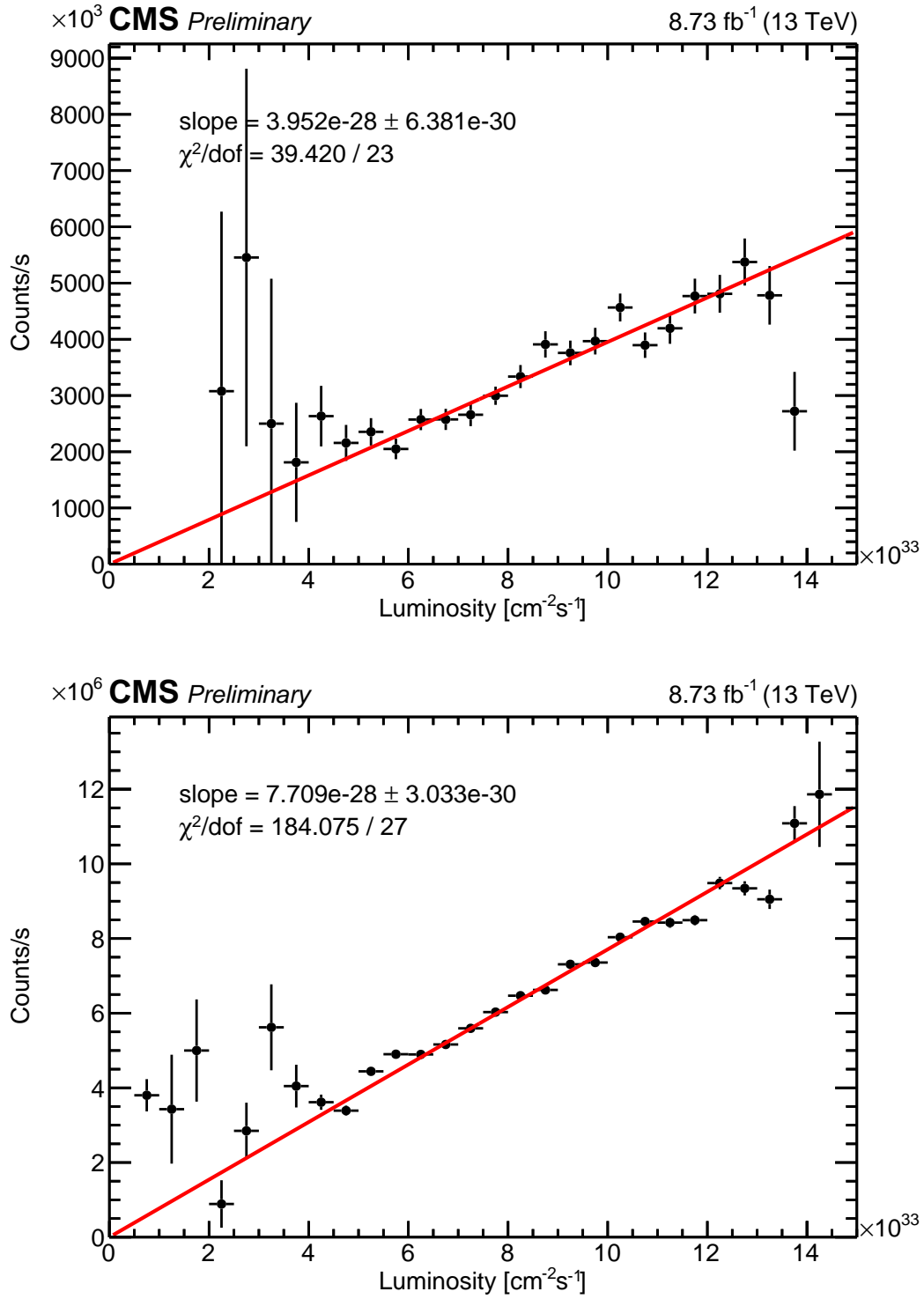


Figure A.25: Plot of ME2/1 wire group  $N_{\text{hits}}^{\text{CMS}}/t$  vs. LHC luminosity for hits that occur (top) during LHC bunch gaps and (bottom) during pp collisions, with a one-parameter linear fit.

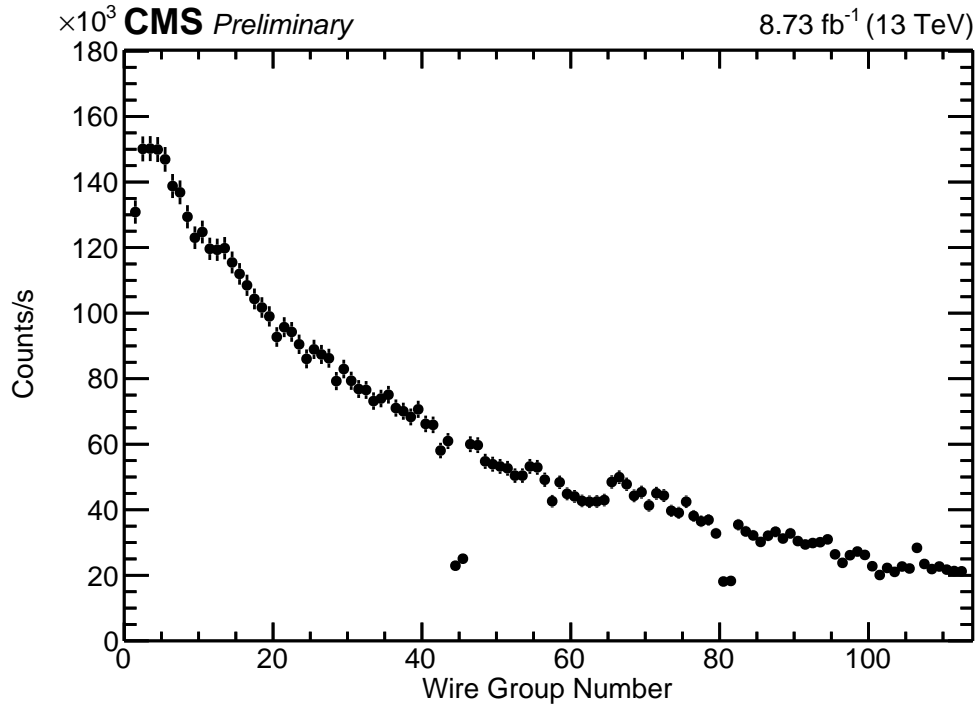


Figure A.26: Plot of anode wire group  $N_{\text{hits}}^{\text{CMS}}(d)/t$  that occur during pp collisions in ME2/1. HV segment 1 is wire group numbers 1–45, HV segment 2 is wire group numbers 46–81, and HV segment 3 is wire group numbers 82–112.

Table A.2: Values of wire group hit fraction and hit rate in each ME2/1 HV segment.

	$f_{\text{ME2/1-}s}$	$H_{\text{WG-ME2/1-}s}$ [hits/s]
ME2/1	-	$6.3 \times 10^6$
ME2/1 S1	0.63	$3.9 \times 10^6$
ME2/1 S2	0.25	$1.5 \times 10^6$
ME2/1 S3	0.12	$0.8 \times 10^6$

We start by calculating the rate of wire group hits in ME2/1 for the full chamber in the same way as ME1/1. Figure A.25 displays plots of the rate of ME2/1 wire group hits second for hits occurring at the end of LHC bunch gaps and during pp collisions with a one-parameter linear fit through the origin. The slopes are again used as the reference hit rate and are combined according to Equation A.25. This gives an estimate of the hit rate in hits/s for ME2/1, denoted as  $H_{\text{WG-ME2/1}}$ .

We then count the number of hits that occur within each HV segment. This number divided by the total number of hits in the chamber is the fraction of hits  $f_{\text{ME2/1-}s}$  that occur in each HV segment, where  $s = 1, 2$ , or  $3$  for each HV segment. The boundaries of HV segments are visible in Figure A.26 as dips in the occupancy plot at wire groups 45 and 81. HV segment 1 is taken as wire group numbers 1–45, HV segment 2 as 46–81, and HV segment 3 as 82–112. Table A.2 reports the hit fractions calculated from Figure A.26. The rate of wire group hits in each HV segment at the reference luminosity is then obtained by multiplying the hit fraction for each HV segment to the full ME2/1 rate of wire group hits,

$$H_{\text{WG-ME2/1-}s} = f_{\text{ME2/1-}s} H_{\text{WG-ME2/1}}. \quad (\text{A.27})$$

These hit rates are also reported in Table A.2.

The next step we take is to convert the luminosity values of all points in Figure A.24 to equivalent hits/s. To do this, we multiply the luminosity value of each point, denoted  $\mathcal{L}$ , by the hits/s at  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ,  $H_{\text{WG-ME2/1-}s}$ , in each HV segment.

$$\text{hits/s} = \mathcal{L} \times H_{\text{WG-ME2/1-s}} \quad (\text{A.28})$$

This transformation from luminosity to equivalent hits/s gives us the HV current as a function of hits/s in each ME2/1 segment individually. Figure A.27 displays example HV current vs. hits/s for ME+2/1/04 Layer 5 and Layer 6.

At CMS, the non-ME1/1 chambers use a different HVPS system than the ME1/1 chambers. When the non-ME1/1 HVPS were commissioned in CMS, each HV channel in the non-ME1/1 chambers was independently calibrated from each other. As a check of this, we study plots of current vs. luminosity and we verify that it is not necessary to correct for fitted offsets when computing the fitted reference current values. However, we observe that for some non-ME1/1 chambers, there are layer-by-layer differences in current in a given chamber and given LHC fill at the same hit rate. Figure A.27 displays example HV current vs. hits/s plots for ME+2/1/04 Layer 5 and Layer 6. The top plot (Layer 5) shows that when different HV segments in the same chamber have similar hits/s, they do have similar HV currents. However, for some channels, a noticeable jump is observed in the HV current at similar hits/s, as in the bottom plot (Layer 6). This discrepancy might be caused by some amount of miscalibration of the HVPS conversion of ADC to  $\mu\text{A}$ . To avoid channels with possible miscalibration, we scan through plots of HV current vs. hits/s for all layers of ME2/1 chambers and search for subjectively defined smooth transitions in HV current and hits/s between each segment.

We then calculate charge/hit by dividing the  $y$  axis value (HV current) by the  $x$  axis value (hits/s) of each data point in the HV current vs. hits/s plots for all three HV segments in each selected chamber layer. Each charge/hit value is collected into a histogram for each LHC fill during run period H of 2016. One entry in this histogram is a single data point from plots of HV current vs. hits/s in chamber layers that pass our selection. Figure A.28 is an example of the charge/hit histogram of selected chamber layers for each HV segment for LHC Fill 5394. (Fill 5394 was used to determine the selection of good ME2/1 layers.) HV segment 1 is the blue histogram, HV segment 2 is the green histogram, and HV segment

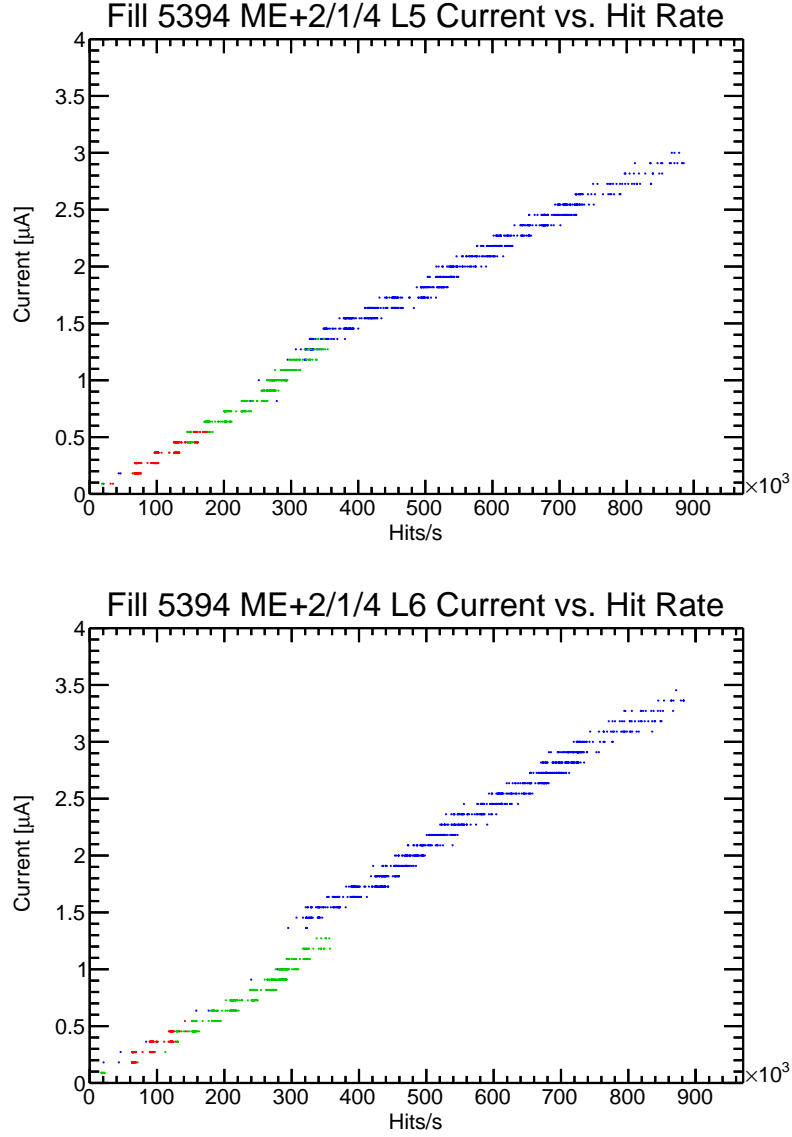


Figure A.27: Plot of HV current vs. hits/s for an ME2/1 layer (top) with smooth transitions in current and hits/s between HV segments and (bottom) Plot of HV current vs. hits/s for an ME2/1 layer with a jump in HV current and hits/s between HV segments 1 and 2. Segment 1 is in blue, segment 2 is in green, and segment 3 is in red.



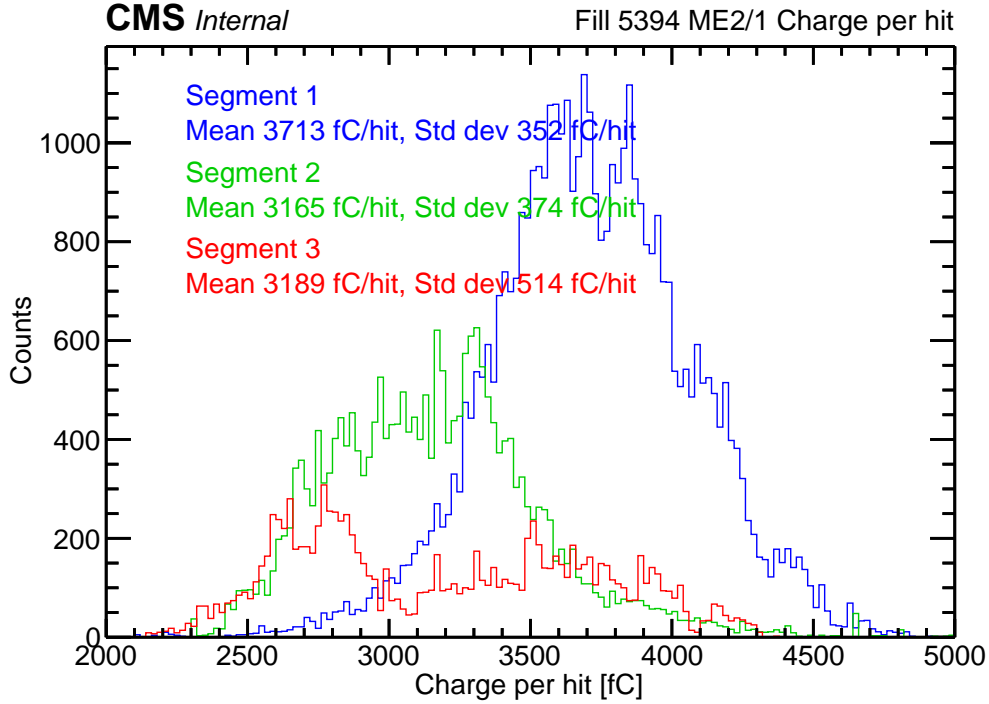


Figure A.28: Histogram of values of charge/hit in fC in ME2/1 for LHC Fill 5394.

3 is the red histogram. HV segment 3 is bi-modal because at very low HV current (less than approximately  $0.6 \mu\text{A}$ ) the HV current ADC least count of approximately  $0.1 \mu\text{A}$  is the dominant source of uncertainty which causes measured values of the HV current to fluctuate between the two closest least count values.

To obtain a charge/hit value in each fill and HV segment, we use the mean of each of the charge/hit histogram. Then finally, just as the HV current measurements in ME1/1 were combined, the charge/hit for each HV segment in each LHC fill is averaged with a weighting according to the integrated luminosity of each fill considered. For HV segment 3 this procedure gives a 4% larger charge/hit than by computing the ratio of the HV current averaged over fills and the hits/s from Table A.2.

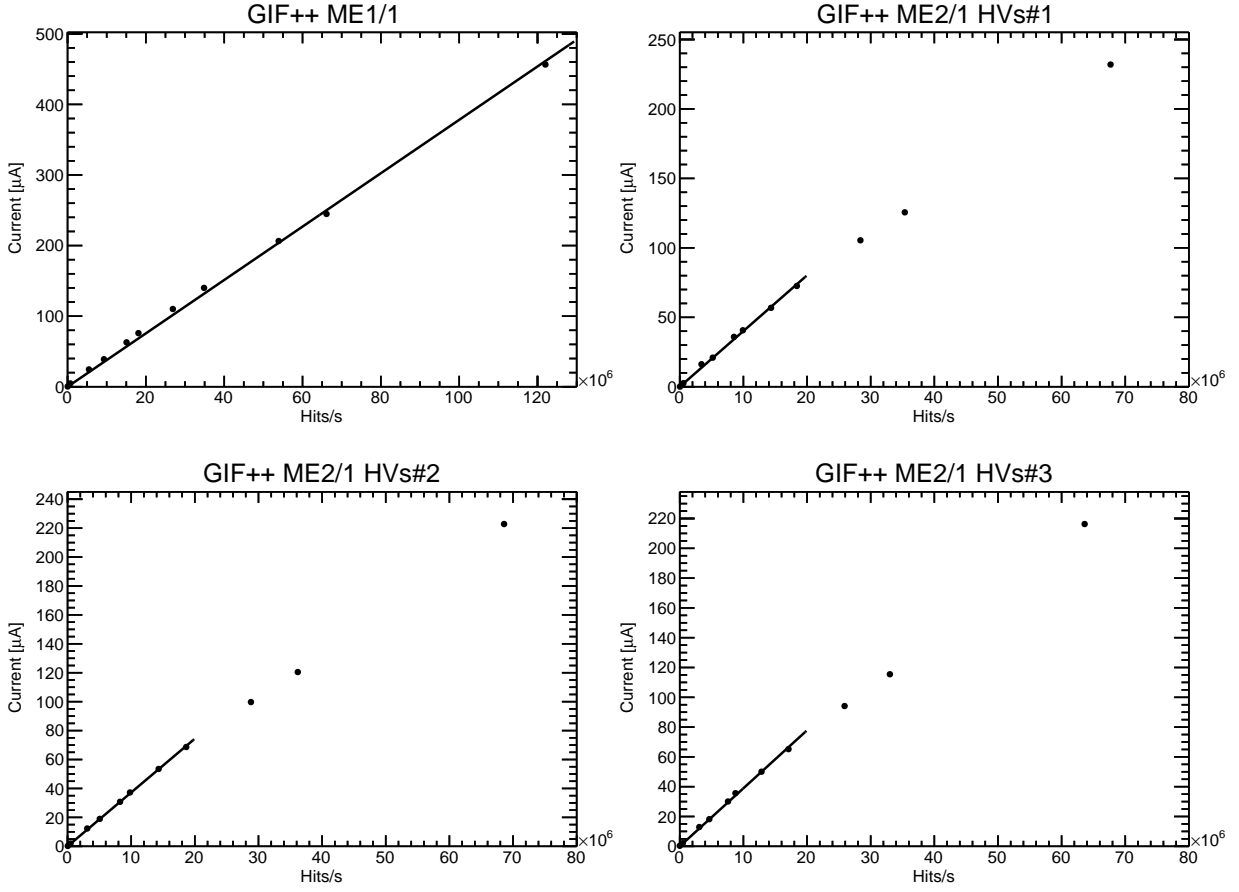


Figure A.29: Plot of anode HV current vs. hits/s at GIF++ for (top left) ME1/1 and ME2/1 HV segment (top right) 1, (bottom left) 2, and (bottom right) 3.

## Computation of charge/hit from GIF++ data

At GIF++, the charge/hit for background hits is obtained by measuring the anode HV current and the rate of early time anode wire groups hits in externally triggered data at each source attenuation. Figure A.29 displays plots of the measured anode HV current as a function of the total number of anode wire group hits that occur in readout time bins 1–5 for the ME1/1 and ME2/1 chambers at GIF++ respectively. Each dot represents a single source attenuation, where the lowest source intensity yields the lowest HV current and lowest background hits/s, and increasing the source intensity also increases the HV current and anode wire hits/s roughly linearly. The slopes of the fitted lines represent the measured charge/hit of each CSC. For ME2/1, only the first 8 points are used to calculate the slope, because at higher source intensities, the high HV currents passing through the internal resistance result in a non-negligible drop in the HV at the anode wire.

## Computation of charge/hit from CMS simulation

In the HP neutron simulation as described above, each simulated wire hit can be associated to a quantity of simulated avalanche charge produced from the gas ionization. Figure A.30 displays histograms of the simulated avalanche charges for ME1/1 and non-ME1/1 simulated wire hit. We take the mean of these histograms of total avalanche charge per simulated wire hit as a measure of charge/hit in MC simulation for neutron-induced hits.

## Comparison of charge/hit

Table A.3 is the collection of all charge/hit calculations in units of fC/hit for ME1/1 and ME2/1 for CMS data, GIF++ data, and CMS simulation.

To summarize the results:

- ME1/1 charge/hit at GIF++ is roughly a factor of 2.5 higher than at CMS
- ME1/1 MC charge/hit matches ME1/1 charge/hit at CMS within 15%

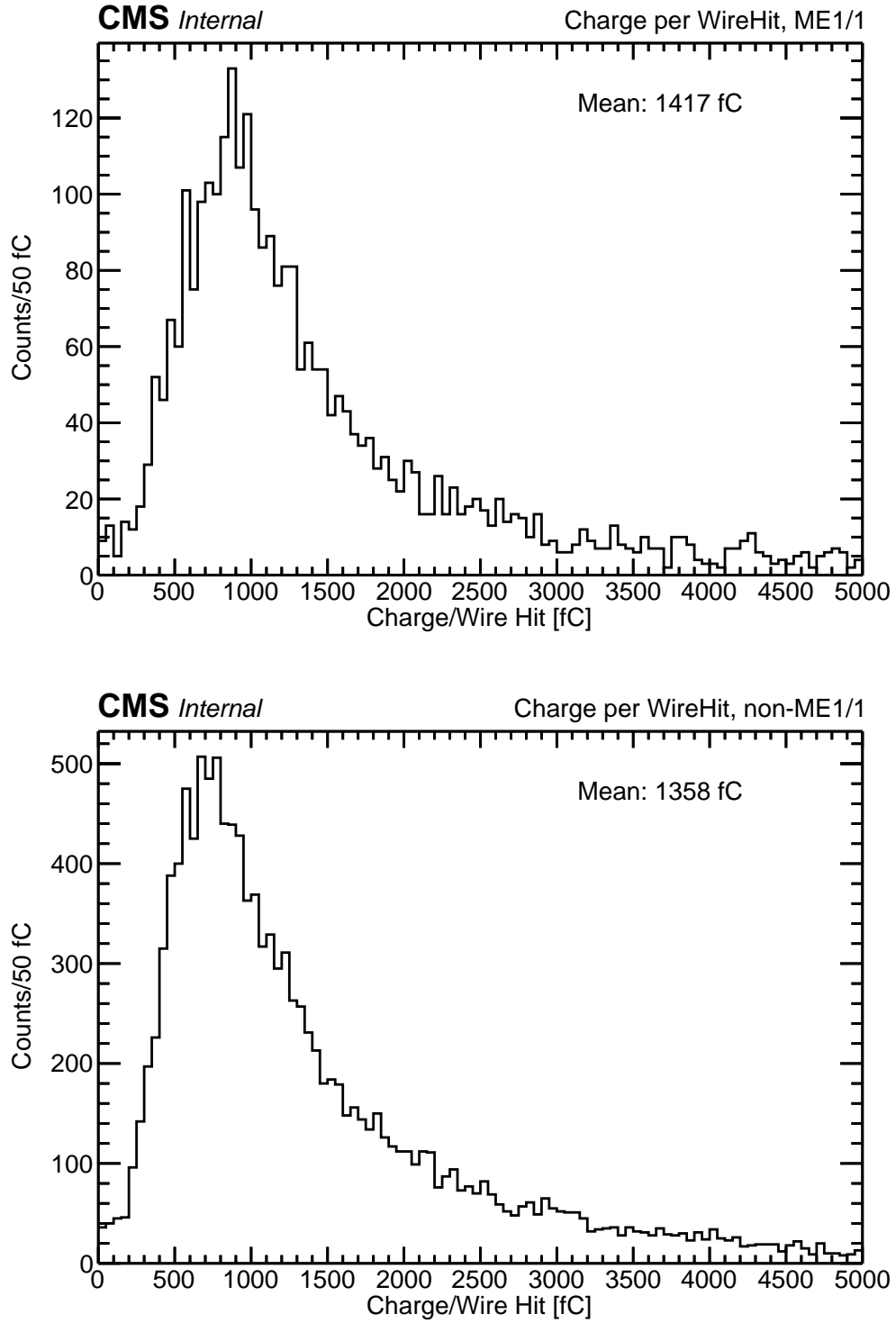


Figure A.30: In simulation of CMS data, histogram of the avalanche charge per simulated wire hit for (top) ME1/1 and (bottom) non-ME1/1 chambers

Table A.3: Values of charge/hit in ME1/1 and each HV segment of ME2/1 in units of fC/hit. All non-ME1/1 chamber charge/hit were observed to be similar. Therefore, the charge/hit was measured by averaging all non-ME1/1 chambers; the \* in the ME2/1 MC charge/hit measurements indicates this.

[fC/hit]	CMS	GIF++	MC
ME1/1	1570	3780	1360
ME2/1 S1	3451	4010	1420*
ME2/1 S2	3274	3720	
ME2/1 S3	3166	3890	

- ME2/1 charge/hit is a factor of 2–3 higher than ME1/1 charge/hit at CMS
- ME2/1 charge/hit at GIF++ is 20–30% higher at GIF++ than at CMS
- ME2/1 MC charge/hit is roughly 2–2.5 lower than ME2/1 charge/hit at CMS

These results are important for reconstruction studies, where the total hits/s is the strongest indicator of trigger and software performance. This means that at equal hits/s for CSCs at GIF++ and CMS (effects on muon reconstruction are roughly the same), the HV current will differ by the same amount as charge/hit. So, when using HV current for correspondence between GIF++ source intensities and ME1/1, these results at face value imply that at an equal hits/s the chamber at GIF++ will have a factor of 2.5 larger HV current than chambers in CMS. Similarly, for the ME2/1 chamber correspondence, the ME2/1 at GIF++ will have a 20–30% larger HV current than the chambers in CMS assuming an equal hits/s (depending on HV segment). However, further understanding of possible miscalibration of HV current ADC to  $\mu\text{A}$  conversion is necessary in the calculation of ME2/1 charge/hit.

## A.8 Acknowledgments of the Appendix

We would like to thank everyone on the CSC GIF++ team, the CMS UCLA group, and everyone whose work we have built on or partially reproduced and who have contributed valuable knowledge and advice, including Cameron Bravo, Tim Cox, Alice Florent, Jay Hauser, Misha Ignatenko, Vladimir Ivantchenko, Evaldas Juska, Alexey Kamenev, Andrey Korytov, Katerina Kuznetsova, Armando Lanaro, Paul Lujan, Alex Madorsky, Nick McColl, Hualin Mei, Yuriy Pakhotin, Vladimir Palchik, Victor Perehygin, Jian Wang, Wells Wulsin, and Piet Verwilligen. This work was partially supported by the U.S. Department of Energy under Award Number DE-SC0009937.

Reference [73] contains brief a history of the measurement of the neutron background in CMS with links to previous work done dating back to 1994.

## REFERENCES

- [1] CMS Collaboration, “Search for long-lived particles that decay into final states containing two electrons or two muons in proton-proton collisions at  $\sqrt{s} = 8$  TeV”, *Phys. Rev. D* **91** (2015) 052012, doi:10.1103/PhysRevD.91.052012.
- [2] CMS Collaboration, “Search for long-lived particles that decay into final states containing two muons, reconstructed using only the CMS muon chambers”, CMS Physics Analysis Summary CMS-PAS-EXO-14-012, CERN, Geneva (2015).
- [3] ATLAS Collaboration, “Search for long-lived particles in final states with displaced dimuon vertices in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”, *Phys. Rev. D* **99** (2019) 012001, doi:10.1103/PhysRevD.99.012001.
- [4] D. J. Griffiths, “Introduction to Elementary Particles, Second Edition”. Wiley-VCH, New York, NY (2008).
- [5] M. Srednicki, “Quantum Field Theory”. Cambridge University Press, New York, NY (2007).
- [6] J. A. Shiflett, “Standard Model Lagrangian”. Retrieved 29 July 2019.  
[http://www.einstein-schrodinger.com/Standard\\_Model.pdf](http://www.einstein-schrodinger.com/Standard_Model.pdf).
- [7] J. Erler and A. Freitas, “Electroweak Model and Constraints on New Physics”, Ch. 10 in Particle Data Group, M. Tanabashi et al., “Review of Particle Physics”, *Phys. Rev. D* **98** (2018) 030001, doi:10.1103/PhysRevD.98.030001.
- [8] M. Carena and others, “Status of Higgs Boson Physics”, Ch. 11 in Particle Data Group, M. Tanabashi et al., “Review of Particle Physics”, *Phys. Rev. D* **98** (2018) 030001, doi:10.1103/PhysRevD.98.030001.
- [9] D. R. Tovey and J. D. Jackson, “Kinematics”, Ch. 47 in Particle Data Group, M. Tanabashi et al., “Review of Particle Physics”, *Phys. Rev. D* **98** (2018) 030001, doi:10.1103/PhysRevD.98.030001.
- [10] M. J. Strassler and K. M. Zurek, “Discovering the Higgs through highly-displaced vertices”, *Phys. Lett. B* **661** (2008) 263, doi:10.1016/j.physletb.2008.02.008.
- [11] D. Curtin, R. Essig, S. Gori, and J. Shelton, “Illuminating dark photons with high-energy colliders”, *JHEP* **02** (2015) 157, doi:10.1007/JHEP02(2015)157.
- [12] The ATLAS Collaboration, The CMS Collaboration, “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at  $\sqrt{s} = 7$  and 8 TeV”, *JHEP* **08** (2016) 45, doi:10.1007/JHEP08(2016)045.
- [13] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

- [14] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [15] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.
- [16] E. Todesco et al., “Training Behavior of the Main Dipoles in the Large Hadron Collider”, *IEEE Trans. Appl. Supercond.* **27** (2017) 1, doi:10.1109/TASC.2017.2657504.
- [17] E. Mobs, “The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018”, Image OPEN-PHO-ACCEL-2018-005, CERN, Geneva (2018). <https://cds.cern.ch/record/2636343>.
- [18] CMS Collaboration, “CMS Web-Based Monitoring: Fill 5423 Bunch Pattern at CMS”. <https://cmswbm.cern.ch/cmsdb/servlet/BunchFill?FILL=5423>.
- [19] S. Baird, “Accelerators for pedestrians; rev. version”, Technical Report CERN-AB-Note-2007-014, CERN-PS-OP-Note-95-17-Rev-2, CERN, Geneva (2007). <http://cds.cern.ch/record/1017689>.
- [20] AC Team, “Diagram of an LHC dipole magnet. Schéma d’un aimant dipôle du LHC”, Image CERN-DI-9906025, CERN, Geneva (1999). <https://cds.cern.ch/record/40524>.
- [21] J.-L. Caron, “Computed magnetic flux in the LHC dipole. Champ magnetique (calcule) dans le dipole LHC.”, Image LHC-PHO-1998-320, CERN, Geneva (1998). <http://cds.cern.ch/record/841503>.
- [22] O. S. Brüning et al., “LHC Design Report”. CERN Yellow Reports: Monographs. CERN, Geneva (2004). <https://cds.cern.ch/record/782076>.
- [23] CMS Collaboration, “CMS Twiki: Public CMS Luminosity Information”. Retrieved 1 May 2019. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [24] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [25] S. R. Davis, “Interactive Slice of the CMS detector”, Image CMS-OUTREACH-2016-027, CERN, Geneva (2016). <http://cds.cern.ch/record/2205172>.
- [26] Y. Hama, “A Note on Lorentz Transformation and Pseudo-Rapidity Distributions”, *Journal of the Phys. Soc. of Japan* **50** (1981) 21, doi:10.1143/JPSJ.50.21.
- [27] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JINST* **13** (2018) P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.



- [28] CMS Collaboration, “The CMS tracker system project: Technical Design Report”, TDR CERN-LHCC-98-006, CMS-TDR-5, CERN, Geneva (1997).  
<https://cds.cern.ch/record/357153>.
- [29] F. Hartmann, “Silicon tracking detectors in high-energy physics”, *Nucl. Instrum. Meth. A* **666** (2012) 25, doi:<https://doi.org/10.1016/j.nima.2011.11.005>.
- [30] CMS Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report”, TDR CERN-LHCC-97-033, CMS-TDR-4, CERN, Geneva (1997).  
<https://cds.cern.ch/record/349375>.
- [31] C. W. Fabjan and F. Gianotti, “Calorimetry for Particle Physics”, *Rev. Mod. Phys.* **75** (2003) 1243, doi:[10.1103/RevModPhys.75.1243](https://doi.org/10.1103/RevModPhys.75.1243).
- [32] CMS Collaboration, “The CMS hadron calorimeter project: Technical Design Report”, TDR CERN-LHCC-97-031, CMS-TDR-2, CERN, Geneva (1997).  
<https://cds.cern.ch/record/357153>.
- [33] A. Penzo and Y. Onel, “The CMS-HF quartz fiber calorimeters”, *Journal of Physics: Conference Series* **160** (2009) 012014, doi:[10.1088/1742-6596/160/1/012014](https://doi.org/10.1088/1742-6596/160/1/012014).
- [34] CMS Collaboration, “The CMS muon project: Technical Design Report”, TDR CERN-LHCC-97-032, CMS-TDR-3, CERN, Geneva (1997).  
<http://cds.cern.ch/record/343814>.
- [35] CMS Collaboration, “The Performance of the CMS Muon Detector in Proton-Proton Collisions at  $\sqrt{s} = 7$  TeV at the LHC”, *JINST* **8** (2013) P11002, doi:[10.1088/1748-0221/8/11/P11002](https://doi.org/10.1088/1748-0221/8/11/P11002), arXiv:[1306.6905](https://arxiv.org/abs/1306.6905).
- [36] CMS Collaboration, “The CMS high level trigger”, *Eur. Phys. J. C* **46** (2006) 605, doi:[10.1140/epjc/s2006-02495-8](https://doi.org/10.1140/epjc/s2006-02495-8), arXiv:[hep-ex/0512077](https://arxiv.org/abs/hep-ex/0512077).
- [37] CMS Collaboration, “CMS Twiki: PdmV 2016 Analysis”. Retrieved 30 May 2019.  
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmV2016Analysis>.
- [38] CMS Collaboration, “CMS Twiki: brilcalc Quick Start”. Retrieved 29 May 2019 using brilcalc 3.5.2.  
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/BrilcalcQuickStart>.
- [39] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, doi:[10.1088/1126-6708/2004/11/040](https://doi.org/10.1088/1126-6708/2004/11/040), arXiv:[hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146).
- [40] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi:[10.1088/1126-6708/2007/11/070](https://doi.org/10.1088/1126-6708/2007/11/070), arXiv:[0709.2092](https://arxiv.org/abs/0709.2092).
- [41] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 043, doi:[10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043), arXiv:[1002.2581](https://arxiv.org/abs/1002.2581).

- [42] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [43] J. Alwall et al., “MadGraph 5: Going Beyond”, *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.
- [44] M. Bähr et al., “Herwig++ physics and manual”, *Eur. Phys. J. C* **58** (2008) 639, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.
- [45] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [46] J. Allison et al., “Geant4 developments and applications”, *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.
- [47] J. Allison et al., “Recent developments in GEANT4”, *Nucl. Instrum. Meth. A* **835** (2016) 186, doi:10.1016/j.nima.2016.06.125.
- [48] C. Liu et al., “Reconstruction of cosmic and beam-halo muons with the CMS detector”, *Eur. Phys. J. C* **56** (2008) 449, doi:10.1140/epjc/s10052-008-0674-7.
- [49] CMS Collaboration, “Search for long-lived particles decaying to final states that include dimuons reconstructed using the muon chambers only”, CMS Analysis Note CMS-AN-2014/176, CERN, Geneva (2014). [http://cms.cern.ch/iCMS/jsp/db\\_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2014/176](http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2014/176).
- [50] CMS Collaboration, “Muon Reconstruction and Identification Improvements for Run-2 and First Results with 2015 Run Data”, CMS Performance Note CMS-DP-2015-015, CERN (2015). <https://cds.cern.ch/record/2037372>.
- [51] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth. A* **262** (1987) 444, doi:10.1016/0168-9002(87)90887-4.
- [52] T. Speer et al., “Vertex Fitting in the CMS Tracker”, CMS Note CMS-NOTE-2006-032, CERN, Geneva (Feb, 2006). <https://cds.cern.ch/record/927395>.
- [53] R. Cousins, J. Linnemann, and J. Tucker, “Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process”, *Nucl. Instrum. Meth. A* **595** (2008) 480, doi:10.1016/j.nima.2008.07.086.
- [54] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, Geneva (2017). <https://cds.cern.ch/record/2257069>.

- [55] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, CMS Note/ATLAS Physics Public Note CMS-NOTE-2011-005/ATL-PHYS-PUB-2011-11, CERN, Geneva (2011). <https://cds.cern.ch/record/1379837>.
- [56] CMS Collaboration, “Documentation for the Higgs Combine Tool”. Retrieved 12 July 2019. <http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/>.
- [57] R. Cousins and V. Highland, “Incorporating systematic uncertainties into an upper limit”, *Nucl. Instrum. Meth. A* **320** (1992) 331, doi:10.1016/0168-9002(92)90794-5.
- [58] R. Cousins, “Probability Density Functions for Positive Nuisance Parameters”. Retrieved 12 July 2019. [http://www.physics.ucla.edu/~cousins/stats/cousins\\_lognormal\\_prior.pdf](http://www.physics.ucla.edu/~cousins/stats/cousins_lognormal_prior.pdf).
- [59] G. Cowan, “Statistics”, Ch. 39 in Particle Data Group, C. Patrignani et al., “Review of Particle Physics”, *Chin. Phys. C* **40** (2016) 100001, doi:10.1088/1674-1137/40/10/100001.
- [60] A. L. Read, “Presentation of search results: The CL<sub>s</sub> technique”, *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.
- [61] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.
- [62] A. Dasgupta, “Study of neutron-induced background hits in the CMS endcap muon system, and implications for the HL-LHC”, poster, The European Physical Society Conference on High Energy Physics, Venice, Italy (5–12 July 2017). <https://indico.cern.ch/event/466934/contributions/2590728/>.
- [63] A. Dasgupta, “Study of neutron-induced background hits in the CMS endcap muon system”, in *Proceedings of The European Physical Society Conference on High Energy Physics*, **314**. PoS(EPS-HEP2017)782. Venice, Italy (2017). doi:10.22323/1.314.0782.
- [64] C. Schnaible, “Study of neutron-induced background hits in the CMS endcap muon system, and implications for the HL-LHC”, contributed talk, APS Division of Particles and Fields Meeting, Fermilab, Batavia, IL, USA (31 July–4 August 2017). <https://indico.fnal.gov/event/11999/session/11/contribution/46>.
- [65] G. Apollinari et al., “High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report”. CERN Yellow Reports: Monographs. CERN, Geneva (2015). <https://cds.cern.ch/record/2116337>.
- [66] CMS Collaboration, “Commissioning of the CMS Experiment and the Cosmic Run at Four Tesla”, *JINST* **5** (2010) T03001, doi:10.1088/1748-0221/5/03/T03001.

- [67] J. Kopecky et al., “Atlas of Neutron Capture Cross Sections”, *International Nuclear Data Committee (IAEA Nuclear Data Section)* **362** (1997).
- [68] J. Hauser et al., “Experience with Trigger Electronics for the CSC System of CMS”, in *Proceedings of the 10th Workshop on Electronics for LHC and Future Experiments*. Boston, USA (2004). doi:10.5170/CERN-2004-010.292.
- [69] D. Acosta et al., “Design features and test results of the CMS endcap muon chambers”, *Nucl. Instrum. Meth. A* **494** (2002) 504, doi:10.1016/S0168-9002(02)01540-1.
- [70] D. Acosta et al., “Efficiency of finding muon track trigger primitives in CMS cathode strip chambers”, *Nucl. Instrum. Meth. A* **592** (2008) 26, doi:10.1016/j.nima.2008.03.118.
- [71] M. Baarmand et al., “Spatial resolution attainable with cathode strip chambers at the trigger level”, *Nucl. Instrum. Meth. A* **425** (1999) 92, doi:10.1016/S0168-9002(98)01390-4.
- [72] T. Sjöstrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.
- [73] P. Verwilligen, “CMS Twiki: Neutron Background in the Muon System”. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonNeutronBackground>.
- [74] R. Cousins, J. Mumford, and S. Valuev, “CSC Trigger Primitive Rates in ORCA”, CMS Note CMS-NOTE-2002-007, CERN, Geneva (2002). <http://cds.cern.ch/record/687399>.
- [75] CMS Collaboration, “Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 13$  TeV”, *JHEP* **07** (2018) 161, doi:10.1007/JHEP07(2018)161.
- [76] CMS Collaboration, “CMS Twiki: Utilities for Accessing Pileup Information for Data”. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData>.
- [77] T. A. Bawej et al., “The New CMS DAQ System for Run 2 of the LHC”, CMS Conference Report CMS-CR-2014-082, CERN, Geneva (2014). <https://cds.cern.ch/record/1711011>.
- [78] D. Pfeiffer et al., “The radiation field in the Gamma Irradiation Facility GIF++ at CERN”, *Nucl. Instrum. Meth. A* **866** (2017) 91, doi:10.1016/j.nima.2017.05.045.
- [79] CMS Collaboration, “Fast algorithm for track segment and hit reconstruction in the CMS Cathode Strip Chambers”, *Nucl. Instrum. Meth. A* **589** (2008) 383, doi:10.1016/j.nima.2008.02.096.